Statistical Interpolation of Spatially Varying but Sparsely Measured 3D Geo-Data Using Compressive Sensing and Variational Bayesian Inference

Tengyuan Zhao[1] · **Yu Wang**[2]

Abstract Real geo-data are three-dimensional (3D) and spatially varied, but measurements are often sparse due to time, resource, and/or technical constraints. In these cases, the quantities of interest at locations where measurements are missing must be interpolated from the available data. Several powerful methods have been developed to address this problem in real-world applications over the past several decades, such as two-point geo-statistical methods (e.g., kriging or Gaussian process regression, GPR) and multiple-point statistics (MPS). However, spatial interpolation remains challenging when the number of measurements is small because a suitable covariance function is difficult to select and the parameters are challenging to estimate from a small number of measurements. Note that a covariance function form and its parameters are key inputs for some methods (e.g., kriging or GPR). MPS is a non-parametric simulation method that combines training images as prior knowledge for sparse measurements. However, the selection of a suitable training image for continuous geo-quantities (e.g., soil or rock properties) faces certain difficulties and may become increasingly complicated when the geo-data to be interpolated are high-dimensional (e.g., 3D) and exhibit non-stationary (e.g., with unknown trends or non-stationary covariance structure) and/or anisotropic characteristics. This paper proposes a non-parametric approach that systematically combines compressive sensing and variational Bayesian inference for statistical interpolation of 3D geo-data. The

✉ Yu Wang
yuwang@cityu.edu.hk

Tengyuan Zhao
tyzhao@xjtu.edu.cn

1   School of Human Settlements and Civil Engineering, Xi'an Jiaotong University, Xi'an, Shaanxi Province, China

2   Department of Architecture and Civil Engineering, City University of Hong Kong, Tat Chee Avenue, Kowloon, Hong Kong

method uses sparse measurements and their locations as the input and provides interpolated values at unsampled locations with quantified interpolation uncertainty as the output. The proposed method is illustrated using a series of numerical 3D examples, and the results indicate a reasonably good performance.

Keywords Spatial data · Bayesian compressive sensing · Data-driven approach · Uncertainty quantification · Non-parametric interpolation

1 Introduction

High-resolution and spatially varied three-dimensional (3D) geo-data play an essential role in geosciences. For instance, high-resolution 3D geophysical data are useful for developing geological models (Wang et al. 2017), which are needed to understand the subsurface conditions in an engineering project (Hillier et al. 2014). High-resolution 3D data of soil or rock properties are a key input for computational models (e.g., 3D finite element models) used to analyze the stability of geo-structures (e.g., slopes, tunnels) (Hack et al. 2006; Griffiths and Marquez 2007; Xiao et al. 2016; Hong et al. 2020). In environmental engineering, high-resolution and spatially varied soil contamination data are necessary for informed and effective decision-making (Largueche 2006; Li and Heap 2011; Horta et al. 2013). High-resolution and spatially varied ore-grade data within a mineral deposit are also a key element in mining engineering (Matheron 1963).

The measurement of high-resolution and spatially varied 3D geo-data is generally expensive and time-consuming because numerous drilled boreholes may be required to obtain adequate subsurface data in 3D space. In geoscience practice, the number of measurements of a given spatially varied quantity of interest is often small due to time, resource, and/or technical constraints. This is especially true for medium- or small-sized projects (Schnabel et al. 2004; Li and Heap 2011). In these cases, the spatially varied quantities of interest at locations where measurements are missing must be interpolated from sparse measurements obtained at other locations, which leads to the classical spatial interpolation problem. Two-point geo-statistical methods, such as kriging or Gaussian process regression (GPR) in machine learning, are commonly used because they provide both the interpolation and uncertainty at unsampled locations (Matheron 1963, 1973; Cressie 1993; Rasmussen and Williams 2006). The uncertainty is a useful indicator that reflects the accuracy and reliability of the interpolation results. After developments in recent decades, GPR and kriging can now handle data with sophisticated two-point covariance structures (Williams 1998; Rasmussen and Williams 2006; Shekaramiz et al. 2019; Gilanifar et al. 2019). Kriging and GPR are powerful in real-world applications but may have some limitations when covariance models of sparse measurements cannot be appropriately determined, partly due to the difficulty in obtaining reliable scale parameter estimates from the very limited number of measurements (Lamorey and Jacobson 1995; Largueche 2006). Additionally, two-point geostatistics cannot handle datasets with complex spatial structures. These limitations have partly motivated the development of multiple-point statistics (MPS) since the early 2000s.

In contrast to kriging or GPR, MPS bypasses the two-point correlation model (e.g., semi-variogram model) (Strebelle 2002) and instead simulates the property of interest at unsampled locations in a non-parametric manner by systematically integrating a training image with measurements (Remy et al. 2009; Mariethoz and Caers 2015). Because of its data-driven nature, MPS has evolved as a non-parametric tool in geological engineering (Zhang 2008), groundwater flow modeling (Huysmans and Dassargues 2009), climate modeling (Mariethoz and Caers 2015), and subsurface layer modeling (Shi and Wang 2020). Notably, MPS requires training images to simulate or interpolate physical quantities at unmeasured locations, and such images can be difficult to obtain, especially for certain continuous variables (e.g., soil/rock properties) over spatial coordinates in 3D space. In addition to MPS, other non-Gaussian non-stationary methods have also been developed to model complex spatial structures in geology, such as the high-order spatial cumulant method (Dimitrakopoulos et al. 2010), which unfortunately requires a relatively large number of data points.

To address this problem and provide an alternative to spatial interpolation, a non-parametric method is developed in this study by combining compressive sensing (CS) from signal processing (Candès and Wakin 2008) and variational Bayesian inference (VBI) from Bayesian inference and machine learning (Bishop 2006). The proposed method requires only a small set of linear measurements and their locations as the input, and provides both the interpolated 3D dataset and the quantified uncertainty associated with the interpolation at each location as the output. The CS framework for 3D geo-data is briefly reviewed in this paper, followed by a detailed description of the Bayesian formulation and VBI and a step-by-step implementation procedure. Finally, the proposed method is illustrated using a series of numerical examples.

2 Framework of Compressive Sensing

In this section, the CS framework is discussed for spatially varied 3D geo-data (e.g., geo-quantity variations in three directions). The notation used in this study is defined before discussing the CS framework in detail. In this study, a bold, underlined, capital letter represents a 3D dataset. For example, $\underline{\mathbf{F}}$ represents a 3D dataset with dimensions of $N_1 \times N_2 \times N_3$. A bold capital letter without an underline represents a two-dimensional (2D) matrix. For example, $\mathbf{B}^1$, $\mathbf{B}^2$, and $\mathbf{B}^3$ represent three 2D matrices with dimensions of $N_1 \times N_1$, $N_2 \times N_2$, and $N_3 \times N_3$, respectively. A bold italic letter is used to represent a vector, such as $\boldsymbol{y}$. Symbols together with a bracket "$(\cdot)$" and indexes are used for an element of a 3D dataset or 2D matrix. For example, $\underline{F}(m, n, l)$ represents an element of $\underline{\mathbf{F}}$ indexed by (m, n, l) and $B^1(m, n)$ represents an element of $\mathbf{B}^1$ indexed by m and n. When an element of a vector is of interest, a vector symbol with a bracketed "$(\cdot)$" or subscripted index is used. For example, $y(m)$ and y_m both represent an element of $\boldsymbol{y}$ indexed by m.

Compressive sensing or sampling is a new sampling strategy used to reconstruct a signal (e.g., 3D dataset) from a small set of linear measurements of that signal. This approach is originated from the field of signal processing (Candès and Romberg 2005; Candès and Wakin 2008) and asserts that most natural signals are compressible or sparse, suggesting that a signal can be represented concisely by a weighted summation
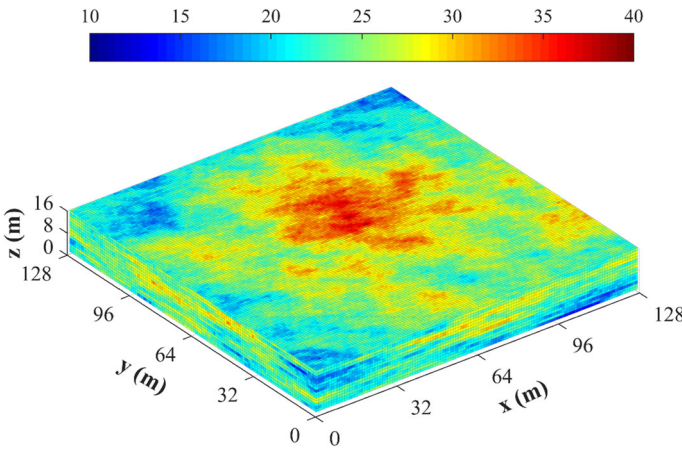
Fig. 1 An illustrative example of 3D dataset $\underline{\mathbf{F}}$

of a limited number of basis functions (e.g., discrete cosine functions) (Salomon 2007; Zhao and Wang 2018a). The signals of interest here may have a spatial variation of mean with magnitude significantly larger than zero. Consider for example, the 3D dataset $\underline{\mathbf{F}}$ in Fig. 1, which has dimensions of $N_1 \times N_2 \times N_3$. Mathematically, $\underline{\mathbf{F}}$ may be represented as a weighted summation of $N = N_1 \times N_2 \times N_3$ 3D basis functions (Caiafa and Cichocki 2013a, b), as illustrated by Fig. 2 and expressed as

$$\underline{\mathbf{F}} = \sum_{t=1}^{N} \omega_t^{3D} \underline{\mathbf{B}}_t^{3D}, \tag{1}$$

where $\underline{\mathbf{B}}_t^{3D}$ is the tth 3D basis function with the same dimension as $\underline{\mathbf{F}}$ and is used to decompose $\underline{\mathbf{F}}$. The construction of $\underline{\mathbf{B}}_t^{3D}$, however, is independent of $\underline{\mathbf{F}}$. A detailed construction of $\underline{\mathbf{B}}_t^{3D}$ is given in "Appendix A". ω_t^{3D} is the weight corresponding to $\underline{\mathbf{B}}_t^{3D}$. For a compressible signal, most elements of ω_t^{3D} are negligible except for a few non-trivial elements with significantly large magnitudes (Caiafa and Cichocki 2013a, b). It is therefore possible to reconstruct the 3D dataset $\underline{\mathbf{F}}$ and estimate or interpolate values of the quantity at unsampled locations by using sparse measurements of $\underline{\mathbf{F}}$, once the non-trivial elements of ω_t^{3D} can be appropriately estimated. In such cases, $\underline{\mathbf{F}}$ can be reconstructed as $\hat{\underline{\mathbf{F}}}$, which is expressed as

$$\hat{\underline{\mathbf{F}}} = \sum_{t=1}^{N} \hat{\omega}_t^{3D} \underline{\mathbf{B}}_t^{3D}, \tag{2}$$

where $\hat{\omega}_t^{3D}$ represents ω_t^{3D} as estimated from measurements.

To estimate $\hat{\omega}_t^{3D}$ from measurements of $\underline{\mathbf{F}}$, a relationship between $\hat{\omega}_t^{3D}$ and the measurements is required. Let a column vector $\boldsymbol{\omega}^{3D}$ represent ω_t^{3D} ($t = 1, 2, ..., N$) and $\boldsymbol{y}^{3D}$ represent M sparse measurements of $\underline{\mathbf{F}}$ with an $M \times 3$ matrix $\mathbf{I}^{\text{ind},M}$ that records the index (m, n, l) of the M elements of $\boldsymbol{y}^{3D}$ in $\underline{\mathbf{F}}$. The first, second, and third
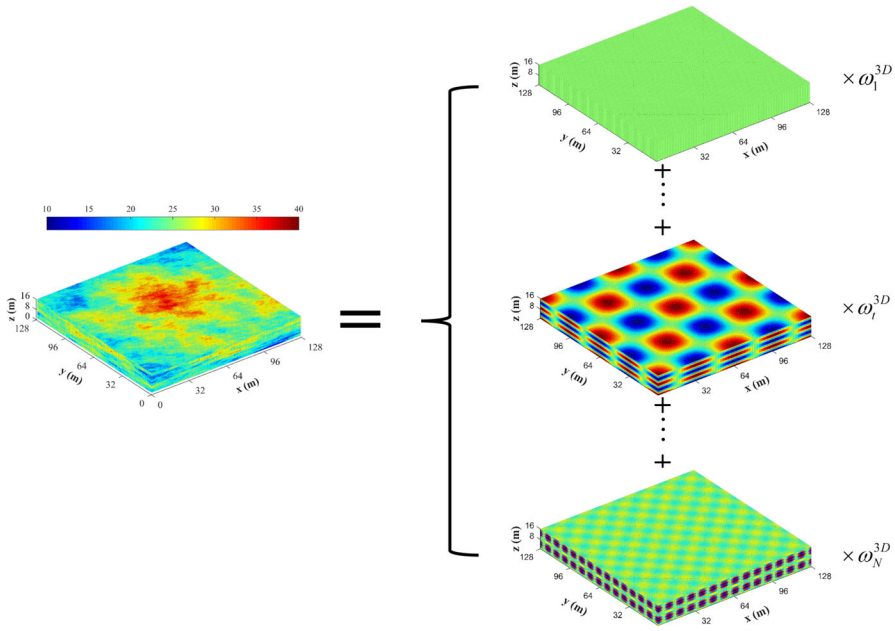
Fig. 2 Illustration of 3D dataset decomposition

columns of $\mathbf{I}^{\text{ind},M}$, i.e., $\boldsymbol{I}_1^{ind,M}$, $\boldsymbol{I}_2^{ind,M}$, and $\boldsymbol{I}_3^{ind,M}$, respectively record the index m, n, and l of the M measurements of $\boldsymbol{y}^{3D}$ in $\underline{\mathbf{F}}$. For mathematical convenience, $\boldsymbol{y}^{3D}$ records the M measurements $\underline{\mathrm{F}}(m, n, l)$ in an increasing order of m ($m \in \boldsymbol{I}_1^{ind,M}$) first, followed by n ($n \in \boldsymbol{I}_2^{ind,\overline{M}}$) and l ($l \in \boldsymbol{I}_3^{ind,\overline{M}}$). A close examination of Eq. (1) shows that each element $\underline{\mathrm{F}}(m, n, l)$ is a weighted summation of $\underline{\mathrm{B}}_t^{3D}(m, n, l)$, i.e., $\underline{\mathrm{F}}(m, n, l) = \sum_{t=1}^{N} \omega_t^{3D} \underline{\mathrm{B}}_t^{3D}(m, n, l)$. For conciseness, $\underline{\mathrm{B}}_t^{3D}(m, n, l)$ ($t = 1, 2, ..., N$) is rewritten as a row vector $(\boldsymbol{a}^r)^{\mathrm{T}}$ with a length of N, where "T" in the superscript represents the transpose operation in linear algebra. In such a case, $\underline{\mathrm{F}}(m, n, l)$ can be rewritten as $(\boldsymbol{a}^r)^{\mathrm{T}} \boldsymbol{\omega}^{3D}$. Because each element of $\boldsymbol{y}^{3D}$ is also an element of $\underline{\mathbf{F}}$, the former can also be expressed as $(\boldsymbol{a}^r)^{\mathrm{T}} \boldsymbol{\omega}^{3D}$. Because there are M elements in $\boldsymbol{y}^{3D}$, there are M row vectors $(\boldsymbol{a}_1^r)^{\mathrm{T}}$, $(\boldsymbol{a}_2^r)^{\mathrm{T}}$, ..., $(\boldsymbol{a}_M^r)^{\mathrm{T}}$, which are stored as a matrix $\mathbf{A}$ with dimensions of $M \times N$. The relationship between $\boldsymbol{y}^{3D}$ and $\boldsymbol{\omega}^{3D}$ is then established as

$$\boldsymbol{y}^{3D} = \mathbf{A}\boldsymbol{\omega}^{3D}. \tag{3}$$

Subsequently, the non-trivial elements of $\boldsymbol{\omega}^{3D}$ may be obtained from $\boldsymbol{y}^{3D}$ using Eq. (3). Because the number of measurements in $\boldsymbol{y}^{3D}$ is usually small, the $\boldsymbol{\omega}^{3D}$ estimated from $\boldsymbol{y}^{3D}$ and denoted as $\hat{\boldsymbol{\omega}}^{3D}$ may contain significant statistical uncertainty. Such uncertainty propagates to the interpolated 3D geo-data $\hat{\underline{\mathbf{F}}}$ through Eq. (2) and may significantly affect the subsequent analysis when the interpolated geo-data are used as the input to analyze problems in geoscience (Luo et al. 2013; Ching et al. 2016; Wang and Zhao 2017b). Quantification of the uncertainty associated with the interpolated
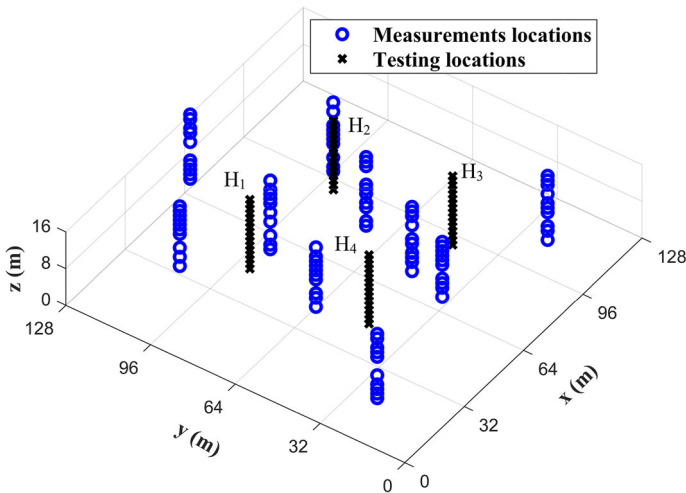
Fig. 3 Locations of measurements and four boreholes H_1 to H_4 in 3D space

3D geo-data $\hat{\underline{\mathbf{F}}}$ is thus essential. This begins with the quantification of the estimation uncertainty associated with $\hat{\omega}^{3D}$, which is discussed in the following section.

Note that in the context of CS, measurements from sparse or compressible signals are often made in a smart way such that each measurement contains valuable information about the entire signal through, say, a linear combination of all measurements in a randomized matrix. Such an operation is often performed for data compression in image compression and signal transmission, where the complete signal or at least a significant portion (e.g., 50%) of the signal is available. However, when only a small portion (e.g., a few percent) of the signal is measured, it might not be necessary to perform a random linear combination for data compression and transmission because the number of data points is already very small. Consider, for example, the data shown in Fig. 3, where only $100/262,144 \approx 0.038\%$ of the dataset is measured. This is a scenario similar to typical geoscience applications (e.g., geological or mining applications), where only a very limited number of soil or rock samples is collected at pre-selected locations (e.g., some prescribed boreholes in Fig. 3), and their properties are measured via either laboratory or in situ tests. In this case, the measurements of the soil and rock properties are similar to a limited number of pixels in an image, rather than "compressed" pixel data. Accordingly, the "sensing matrix" is only a matrix with all elements being either one or zero, which reflects the measurement positions of the soil/rock properties in space. A random linear combination is therefore not used in this study.

3 Bayesian Formulation for $\hat{\omega}^{3D}$

In this section, $\hat{\omega}^{3D}$ is estimated under a Bayesian framework, and its uncertainty is effectively quantified. A Bayesian method is adopted in this study because it can effectively handle various uncertainties, including model uncertainty (Zhang et al.

2009; Wang and Zhao 2017a), spatial variability of soil properties (Wang et al. 2016; Wang and Zhao 2017b), and soil stratification or lithology characterization uncertainty (Buland et al. 2008; Wang et al. 2013, 2018; Cao et al. 2018; Moja et al. 2018). Following Bayes' theorem, the knowledge (including uncertainty) of $\hat{\boldsymbol{\omega}}^{3D}$ updated by measurements $\boldsymbol{y}^{3D}$ is reflected by its posterior probability density function (PDF), i.e., $p(\hat{\boldsymbol{\omega}}^{3D}|\boldsymbol{y}^{3D}, Prior)$, where "$Prior$" represents the knowledge of $\hat{\boldsymbol{\omega}}^{3D}$ in the absence of measurements $\boldsymbol{y}^{3D}$. In Bayesian literature, $p(\hat{\boldsymbol{\omega}}^{3D}|\boldsymbol{y}^{3D}, Prior)$ is usually simplified as $p(\hat{\boldsymbol{\omega}}^{3D}|\boldsymbol{y}^{3D})$, which is expressed as (Sivia and Skilling 2006)

$$p(\hat{\boldsymbol{\omega}}^{3D}|\boldsymbol{y}^{3D}) = \frac{p(\boldsymbol{y}^{3D}|\hat{\boldsymbol{\omega}}^{3D})p(\hat{\boldsymbol{\omega}}^{3D})}{p(\boldsymbol{y}^{3D})}, \tag{4}$$

where $p(\boldsymbol{y}^{3D}|\hat{\boldsymbol{\omega}}^{3D})$ is the likelihood function, which reflects the plausibility of observing $\boldsymbol{y}^{3D}$ given $\hat{\boldsymbol{\omega}}^{3D}$; $p(\hat{\boldsymbol{\omega}}^{3D})$ is the prior PDF of $\hat{\boldsymbol{\omega}}^{3D}$, which reflects the prior knowledge of $\hat{\boldsymbol{\omega}}^{3D}$ in the absence of $\boldsymbol{y}^{3D}$; and $p(\boldsymbol{y}^{3D})$ is a constant, which ensures that the integration of $p(\hat{\boldsymbol{\omega}}^{3D}|\boldsymbol{y}^{3D}) = 1$. Equation (4) shows that the likelihood function and the prior PDF are two essential ingredients of a Bayesian framework; these are constructed separately below.

To construct the likelihood function $p(\boldsymbol{y}^{3D}|\hat{\boldsymbol{\omega}}^{3D})$, the elements of $\boldsymbol{y}^{3D}$ should be related to $\hat{\boldsymbol{\omega}}^{3D}$ [see Eq. (3)]. When $\hat{\boldsymbol{\omega}}^{3D}$ is estimated and substituted into Eq. (3), some residuals may be introduced between $\boldsymbol{y}^{3D}$ and $\mathbf{A}\hat{\boldsymbol{\omega}}^{3D}$. The residuals are often modeled as a Gaussian random variable ε in the literature (Bishop and Tipping 2000; Tipping 2001). In such cases, Eq. (3) is modified to

$$\boldsymbol{y}^{3D} = \mathbf{A}\hat{\boldsymbol{\omega}}^{3D} + \varepsilon I_M, \tag{5}$$

where ε has a mean of zero and an unknown variance. $\boldsymbol{I}_M$ is a column vector with a length of M, and all of its elements are equal to one. The likelihood function of observing $\boldsymbol{y}^{3D}$ is then expressed as (Tipping 2001; Ji et al. 2008)

$$p(\boldsymbol{y}^{3D}|\hat{\boldsymbol{\omega}}^{3D}, \tau) = \frac{\tau^{M/2}}{\left(\sqrt{2\pi}\right)^M} \exp\left(-\frac{\tau\left(\boldsymbol{y}^{3D} - \mathbf{A}\hat{\boldsymbol{\omega}}^{3D}\right)^T\left(\boldsymbol{y}^{3D} - \mathbf{A}\hat{\boldsymbol{\omega}}^{3D}\right)}{2}\right), \tag{6}$$

where τ represents the reciprocal of the unknown variance of ε for derivation convenience. Because τ influences $\hat{\boldsymbol{\omega}}^{3D}$ through Eqs. (6) and (4), τ may also be taken as a random variable (Huang et al. 2016; Bishop and Tipping 2000). In such cases, a prior joint PDF of $(\hat{\boldsymbol{\omega}}^{3D}, \tau)$ is required and can be expressed as

$$p(\hat{\boldsymbol{\omega}}^{3D}, \tau) = p(\hat{\boldsymbol{\omega}}^{3D})p(\tau), \tag{7}$$

where $\hat{\boldsymbol{\omega}}^{3D}$ and τ are taken to be independent of each other.

Consider first the prior PDF of $\hat{\boldsymbol{\omega}}^{3D}$. Each element of $\hat{\boldsymbol{\omega}}^{3D}$, i.e., $\hat{\omega}_i^{3D}$, is taken to follow a Gaussian distribution with a mean of zero and an unknown variance. This may be justified by noting that $\hat{\omega}_i^{3D}$ can be either negative or positive. For derivation

convenience, the inverse of the variance of $\hat{\omega}_t^{3D}$ is expressed as α_t ($\alpha_t > 0$) (Bishop and Tipping 2000; Zhao et al. 2018a). The prior distribution of $\hat{\omega}^{3D}$ is then expressed as

$$p(\hat{\omega}^{3D}|\boldsymbol{\alpha}) = \prod_{t=1}^{N} \frac{\alpha_t^{1/2}}{\sqrt{2\pi}} \exp\left(-\frac{\alpha_t(\omega_t^{3D})^2}{2}\right) = \frac{[\det(\mathbf{D}^\alpha)]^{1/2}}{(2\pi)^{N/2}} \exp\left(-\frac{(\omega^{3D})^T \mathbf{D}^\alpha (\omega^{3D})}{2}\right).$$
(8)

Note that Eq. (8) and the likelihood function in Eq. (6) are a conjugate pair in the Bayesian formulation, which means that the posterior PDF of $\hat{\omega}^{3D}$ to be derived in the next section also follows a multivariate normal distribution (Murphy 2012). In Eq. (8), $\boldsymbol{\alpha}$ represents α_t in a vector format and $\mathbf{D}^\alpha$ is a diagonal matrix with a dimension of $N \times N$, which has diagonal element $D^\alpha(t, t) = \alpha_t$. α_t ($t = 1, 2, \ldots, N$) may be treated probabilistically as a random variable such that the uncertainty associated with α_t and its effect on $\hat{\omega}^{3D}$ can be explicitly considered. Following Zhao et al. (2015), α_t is taken to follow an inverse Gamma distribution with a constant parameter of 1 and γ ($\gamma > 0$), i.e., $p(\alpha_t|\gamma) = \gamma/2 \cdot \alpha_t^{-2} \cdot \exp(-\gamma/2 \cdot \alpha_t^{-1})$. The prior PDF for $\boldsymbol{\alpha}$ is expressed as

$$p(\boldsymbol{\alpha}|\gamma) = \prod_{t=1}^{N} p(\alpha_t|\gamma).$$
(9)

γ is then further taken as a random variable, which follows a Gamma distribution

$$p(\gamma) = \text{Gamma}(a_0, b_0) = b_0^{a_0} \gamma^{a_0-1} \exp(-b_0\gamma) / \Gamma(a_0),$$
(10)

where a_0 and b_0 are taken as small non-negative values in this study, e.g., $a_0 = b_0 = 10^{-4}$. Small a_0 and b_0 values in a Gamma distribution ensure that γ can potentially vary across a very broad range but tends to have small values. In such a case, the prior PDF of α_t tends to exhibit a spike around zero; namely, α_t tends to be small and $1/\alpha_t$, i.e., the variance of $\hat{\omega}_t^{3D}$, tends to be significant. Such parameter configurations ensure that the prior PDF of $\hat{\omega}_t^{3D}$ is uninformative and that $\hat{\omega}_t^{3D}$ can be efficiently updated by the measurements. The reciprocal of the variance of ε, i.e., τ, is taken to follow a Gamma distribution, and the prior PDF $p(\tau)$ is expressed as

$$p(\tau) = \frac{d_0^{c_0} \tau^{c_0-1}}{\Gamma(c_0)} \exp(-d_0\tau),$$
(11)

where c_0 and d_0 are also non-negative parameters and are taken as small values, e.g., $c_0 = d_0 = 10^{-4}$ in this study (Shekaramiz et al. 2017, 2019). This ensures that τ has a tendency of being small, i.e., large values for $1/\tau$. Such a configuration promotes a non-informative prior distribution for the variance of the residuals between y^{3D} and $\mathbf{A}\hat{\omega}^{3D}$.

The joint prior PDF of $\hat{\omega}^{3D}$, $\boldsymbol{\alpha}$, τ, γ can then be obtained by substituting the prior PDF of each parameter shown in Eqs. (8)–(11) into Eq. (12) below

$$p(\hat{\omega}^{3D}, \boldsymbol{\alpha}, \gamma, \tau) = p(\hat{\omega}^{3D}|\boldsymbol{\alpha})p(\boldsymbol{\alpha}|\gamma)p(\gamma)p(\tau).$$
(12)

Note that Eq. (12) uses $p(\hat{\boldsymbol{\omega}}^{3D}, \boldsymbol{\alpha}, \gamma, \tau)$, rather than $p(\hat{\boldsymbol{\omega}}^{3D})$ as in Eq. (7), because $\boldsymbol{\alpha}$, γ, and τ are also taken as random variables in addition to $\hat{\boldsymbol{\omega}}^{3D}$. Subsequently, the joint posterior PDF of $(\hat{\boldsymbol{\omega}}^{3D}, \boldsymbol{\alpha}, \gamma, \tau)$ can be derived theoretically as

$$
\begin{aligned}
p(\hat{\boldsymbol{\omega}}^{3D}, \boldsymbol{\alpha}, \gamma, \tau | \boldsymbol{y}^{3D}) &= p(\boldsymbol{y}^{3D}|\hat{\boldsymbol{\omega}}^{3D}, \tau) p(\hat{\boldsymbol{\omega}}^{3D}|\boldsymbol{\alpha}) p(\boldsymbol{\alpha}|\gamma) p(\gamma) p(\tau) / p(\boldsymbol{y}^{3D}). \\
&= \frac{\tau^{M/2}}{(2\pi)^{M/2}} \exp\left(-\frac{\tau(\boldsymbol{y}^{3D} - \mathbf{A}\hat{\boldsymbol{\omega}}^{3D})^T (\boldsymbol{y}^{3D} - \mathbf{A}\hat{\boldsymbol{\omega}}^{3D})}{2} \right) \\
&\quad \frac{[\det(\mathbf{D}^\alpha)]^{1/2}}{(2\pi)^{N/2}} \exp\left(-\frac{(\hat{\boldsymbol{\omega}}^{3D})^T \mathbf{D}^\alpha (\hat{\boldsymbol{\omega}}^{3D})}{2} \right) \\
&\quad \times \prod_{t=1}^{N} \frac{\gamma}{2} \alpha_t^{-2} \exp\left[-\frac{\gamma}{2} \alpha_t^{-1} \right] \times \frac{d_0^{c_0} \tau^{c_0-1}}{\Gamma(c_0)} \exp(-d_0 \tau) \times 1/p(\boldsymbol{y}^{3D}).
\end{aligned}
$$
(13)

The marginal posterior PDF of $\hat{\boldsymbol{\omega}}^{3D}$ can be obtained theoretically by marginalization, i.e., $p(\hat{\boldsymbol{\omega}}^{3D}|\boldsymbol{y}^{3D}) = \int p(\hat{\boldsymbol{\omega}}^{3D}, \boldsymbol{\alpha}, \gamma, \tau | \boldsymbol{y}^{3D}) d\boldsymbol{\alpha} d\gamma d\tau$. In this case, the influence of the uncertainty of hyperparameters $\boldsymbol{\alpha}$, τ, and γ on $\hat{\boldsymbol{\omega}}^{3D}$ is explicitly considered in the formulation. The 3D dataset $\hat{\mathbf{F}}$ can then be interpolated, and the uncertainty propagated from $\hat{\boldsymbol{\omega}}^{3D}$ can be quantified. However, no analytical solution is available for the joint posterior PDF of $\hat{\boldsymbol{\omega}}^{3D}, \boldsymbol{\alpha}, \gamma, \tau$ because there is no analytical expression for $p(\boldsymbol{y}^{3D})$ (Bishop and Tipping 2000). To obtain the marginal posterior PDF of $\hat{\boldsymbol{\omega}}^{3D}$, the VBI method is adopted, as discussed in the following section.

4 Variational Bayesian Inference (VBI)

The VBI aims to find a tractable distribution that can properly approximate the true joint posterior PDF of interest (Beal 2003). In the context of this study, the VBI aims to find a joint PDF $q(\hat{\boldsymbol{\omega}}^{3D}, \boldsymbol{\alpha}, \gamma, \tau)$ that can approximate the posterior PDF $p(\hat{\boldsymbol{\omega}}^{3D}, \boldsymbol{\alpha}, \gamma, \tau | \boldsymbol{y}^{3D})$. Note that the approximate PDF obtained from the VBI is denoted as "$q(\cdot)$" to distinguish from the true PDF "$p(\cdot)$". In the VBI, the approximate PDF $q(\hat{\boldsymbol{\omega}}^{3D}, \boldsymbol{\alpha}, \gamma, \tau)$ is obtained by minimizing a difference metric, such as Kullback–Leibler (KL) divergence, between $q(\hat{\boldsymbol{\omega}}^{3D}, \boldsymbol{\alpha}, \gamma, \tau)$ and $p(\hat{\boldsymbol{\omega}}^{3D}, \boldsymbol{\alpha}, \gamma, \tau | \boldsymbol{y}^{3D})$. The KL divergence between $q(\hat{\boldsymbol{\omega}}^{3D}, \boldsymbol{\alpha}, \gamma, \tau)$ and $p(\hat{\boldsymbol{\omega}}^{3D}, \boldsymbol{\alpha}, \gamma, \tau | \boldsymbol{y}^{3D})$, denoted as KL$(q\|p)$, is expressed as (Bishop 2006; Murphy 2012; Yu et al. 2016)

$$
KL(q||p) = -\int q(\hat{\boldsymbol{\omega}}^{3D}, \boldsymbol{\alpha}, \gamma, \tau) \ln \frac{p(\hat{\boldsymbol{\omega}}^{3D}, \boldsymbol{\alpha}, \gamma, \tau | \boldsymbol{y}^{3D})}{q(\hat{\boldsymbol{\omega}}^{3D}, \boldsymbol{\alpha}, \gamma, \tau)} d\hat{\boldsymbol{\omega}}^{3D} d\boldsymbol{\alpha} d\gamma d\tau.
$$
(14)

Note that KL$(q\|p)$ is non-negative. Smaller KL$(q\|p)$ values allow $q(\hat{\boldsymbol{\omega}}^{3D}, \boldsymbol{\alpha}, \gamma, \tau)$ to better approximate $p(\hat{\boldsymbol{\omega}}^{3D}, \boldsymbol{\alpha}, \gamma, \tau | \boldsymbol{y}^{3D})$. By reformulating the integrand in Eqs. (14) and (15) is obtained as

$$
ln[p(\boldsymbol{y}^{3D})] = L(q) + \text{KL}(q||p),
$$
(15)

where $\ln[p(\mathbf{y}^{3D})]$ is the natural logarithm of the constant term $p(\mathbf{y}^{3D})$ from Eqs. (4) and (13). Because $p(\mathbf{y}^{3D})$ is a constant, $\ln[p(\mathbf{y}^{3D})]$ is also a constant. $L(q)$ represents the lower bound of $\ln[p(\mathbf{y}^{3D})]$. A detailed derivation of Eq. (15) and $L(q)$ is given in "Appendix B". Equation (15) shows that the minimization of KL$(q\|p)$ is equivalent to the maximization of $L(q)$.

To obtain a tractable distribution, $q(\hat{\boldsymbol{\omega}}^{3D}, \boldsymbol{\alpha}, \gamma, \tau)$ is often factorized as (Bishop 2006)

$$q(\hat{\boldsymbol{\omega}}^{3D}, \boldsymbol{\alpha}, \gamma, \tau) = q(\hat{\boldsymbol{\omega}}^{3D})q(\boldsymbol{\alpha})q(\gamma)q(\tau), \tag{16}$$

where $q(\hat{\boldsymbol{\omega}}^{3D})$, $q(\boldsymbol{\alpha})$, $q(\gamma)$, and $q(\tau)$ respectively represent the tractable PDFs of $\hat{\boldsymbol{\omega}}^{3D}$, $\boldsymbol{\alpha}$, γ, and τ from the VBI, which respectively approximate the true posterior PDF of $\hat{\boldsymbol{\omega}}^{3D}$, $\boldsymbol{\alpha}$, γ, and τ. Following the factorization in Eq. (16) and derivation discussed in Bishop (2006), the approximate PDF of a random variable of interest (e.g., $q(\hat{\boldsymbol{\omega}}^{3D})$) or its logarithmic version (e.g., $\ln[q(\hat{\boldsymbol{\omega}}^{3D})]$) that maximizes $L(q)$ can be expressed in terms of the expectation of the right joint PDF (e.g., $\ln p(\hat{\boldsymbol{\omega}}^{3D}, \boldsymbol{\alpha}, \gamma, \tau, \mathbf{y}^{3D})$) under the approximate posterior PDF of other random variables (e.g., $q(\boldsymbol{\alpha})$, $q(\gamma)$, $q(\tau)$). For example, the $q(\hat{\boldsymbol{\omega}}^{3D})$ that maximizes $L(q)$ is expressed as (see "Appendix C")

$$\ln[q(\hat{\boldsymbol{\omega}}^{3D})] = \int q(\boldsymbol{\alpha})q(\gamma)q(\tau) \ln p(\hat{\boldsymbol{\omega}}^{3D}, \boldsymbol{\alpha}, \gamma, \tau, \mathbf{y}^{3D})\mathrm{d}\boldsymbol{\alpha}\mathrm{d}\gamma\mathrm{d}\tau + const_1, \tag{17}$$

where "$const_1$" is a constant that ensures that the integration of $q(\hat{\boldsymbol{\omega}}^{3D}) = 1$. Based on the conditional probability rules and Eq. (12), $\ln p(\hat{\boldsymbol{\omega}}^{3D}, \boldsymbol{\alpha}, \gamma, \tau, \mathbf{y}^{3D})$ is expressed as

$$\ln p(\hat{\boldsymbol{\omega}}^{3D}, \boldsymbol{\alpha}, \gamma, \tau, \mathbf{y}^{3D}) = \ln p(\mathbf{y}^{3D}|\hat{\boldsymbol{\omega}}^{3D}, \tau) + \ln p(\hat{\boldsymbol{\omega}}^{3D}|\boldsymbol{\alpha}) + \ln p(\boldsymbol{\alpha}|\gamma) + \ln p(\gamma) + \ln p(\tau). \tag{18}$$

Substituting Eqs. (6), (8–11), and (18) into Eq. (17) and rearranging the terms lead to

$$q(\hat{\boldsymbol{\omega}}^{3D}) = \frac{1}{\sqrt{(2\pi)^N \det(\boldsymbol{\Sigma}_{\hat{\boldsymbol{\omega}}^{3D}})}} \exp\left[-\frac{(\hat{\boldsymbol{\omega}}^{3D} - \boldsymbol{\mu}_{\hat{\boldsymbol{\omega}}^{3D}})^T (\boldsymbol{\Sigma}_{\hat{\boldsymbol{\omega}}^{3D}})^{-1}(\hat{\boldsymbol{\omega}}^{3D} - \boldsymbol{\mu}_{\hat{\boldsymbol{\omega}}^{3D}})}{2}\right]. \tag{19}$$

Equation (19) shows that $q(\hat{\boldsymbol{\omega}}^{3D})$ is a multivariate normal distribution, the detailed derivation of which is summarized in "Appendix C". The mean $\boldsymbol{\mu}_{\hat{\boldsymbol{\omega}}^{3D}}$ and covariance matrix $\boldsymbol{\Sigma}_{\hat{\boldsymbol{\omega}}^{3D}}$ are expressed as

$$\begin{aligned}\boldsymbol{\mu}_{\hat{\boldsymbol{\omega}}^{3D}} &= \boldsymbol{\Sigma}_{\hat{\boldsymbol{\omega}}^{3D}}\mathbf{A}^T \mathbf{y}^{3D} E(\tau) \\ \boldsymbol{\Sigma}_{\hat{\boldsymbol{\omega}}^{3D}} &= [\mathbf{A}^T \mathbf{A} E(\tau) + E(\mathbf{D}^{\alpha})]^{-1}, \end{aligned} \tag{20}$$

where $E(\tau) = \int \tau q(\tau)\tau$ represents the posterior mean of τ. Similarly, $E(\mathbf{D}^{\alpha})$ represents the posterior mean of α_t in matrix format, i.e., $E(\mathrm{D}^{\alpha}(t,t)) = E(\alpha_t) = \int \alpha_t q(\alpha_t)\alpha_t$, where $q(\alpha_t)$ represents the distribution that can properly approximate the true posterior PDF

of α_t. Following a derivation procedure similar to that of $q(\hat{\omega}^{3D})$, $q(\alpha_t)$ and $q(\tau)$ are derived to follow a Generalized Inverse Gaussian (GIG) distribution and a Gamma distribution, respectively, as shown in "Appendix C". $E(\alpha_t)$ and $E(\tau)$ are then obtained as

$$E(\alpha_t) = \frac{\sqrt{b_t} K_{p+1}(\sqrt{a_t b_t})}{\sqrt{a_t} K_p(\sqrt{a_t b_t})}, \tag{21a}$$

$$E(\tau) = \frac{c_n}{d_n}, \tag{21b}$$

where $a_t = \mu^2_{\hat{\omega}^{3D}_t} + \sigma^2_{\hat{\omega}^{3D}_t}$; $\mu_{\hat{\omega}^{3D}_t}$ and $\sigma^2_{\hat{\omega}^{3D}_t}$ represent the tth element of $\boldsymbol{\mu}_{\hat{\omega}^{3D}}$ and tth diagonal element of $\boldsymbol{\Sigma}_{\hat{\omega}^{3D}}$, respectively; $b_t = E(\gamma)$ is the posterior mean of γ, i.e., $E(\gamma) = \int \gamma q(\gamma) d\gamma$, which is shown later in this section; $K_{p+1}(\cdot)$ represents the modified Bessel function of the second kind with parameter $p = -1/2$ in this study; and c_n and d_n are the shape parameters of $q(\tau)$, which are expressed as $c_n = M/2 + c_0$ and $d_n = d_0 + 1/2[(\boldsymbol{y}^{3D})^T (\boldsymbol{y}^{3D}) - 2\boldsymbol{\mu}^T_{\hat{\omega}^{3D}} \mathbf{A}^T \boldsymbol{y}^{3D} + E[(\hat{\omega}^{3D})^T \mathbf{A}^T \mathbf{A}(\hat{\omega}^{3D})]]$, respectively. $E[(\hat{\omega}^{3D})^T \mathbf{A}^T \mathbf{A}(\hat{\omega}^{3D})]$ is expressed as (Petersen 2004)

$$E[(\hat{\omega}^{3D})^T \mathbf{A}^T \mathbf{A}(\hat{\omega}^{3D})] = \boldsymbol{\mu}^T_{\hat{\omega}^{3D}} \mathbf{A}^T \mathbf{A} \boldsymbol{\mu}_{\hat{\omega}^{3D}} + T_r(\mathbf{A}\boldsymbol{\Sigma}_{\hat{\omega}^{3D}} \mathbf{A}^T), \tag{22}$$

where "$T_r(\cdot)$" is the trace operation, which represents the sum of all diagonal elements of a matrix. In addition, $q(\gamma)$ is derived to follow a Gamma distribution (see "Appendix C") with the expectation of $E(\gamma)$

$$E(\gamma) = \gamma_a / \gamma_b, \tag{23}$$

where $\gamma_a = N + a_0$ and $\gamma_b = b_0 + \sum_{t=1}^{N} E(\alpha_t^{-1})$. $E(\alpha_t^{-1})$ represents the posterior mean of α_t^{-1}, i.e., $E(\alpha_t^{-1}) = \int \alpha_t^{-1} q(\alpha_t^{-1}) d\alpha_t^{-1}$, and $q(\alpha_t^{-1})$ is derived to follow a GIG distribution (see "Appendix C") with parameters $-p = 1/2$, b_t, and a_t. $E(\alpha_t^{-1})$ is expressed as

$$E(\alpha_t^{-1}) = \frac{\sqrt{a_t} K_{-p+1}(\sqrt{a_t b_t})}{\sqrt{b_t} K_{-p}(\sqrt{a_t b_t})}. \tag{24}$$

The above derivations show that $\boldsymbol{\mu}_{\hat{\omega}^{3D}}$ and $\boldsymbol{\Sigma}_{\hat{\omega}^{3D}}$ depend on $E(\tau)$ and $E(\alpha_t)$, which in turn depend on $\boldsymbol{\mu}_{\hat{\omega}^{3D}}$, $\boldsymbol{\Sigma}_{\hat{\omega}^{3D}}$, and $E(\gamma)$. As a result, an iteration process among $\boldsymbol{\mu}_{\hat{\omega}^{3D}}$, $\boldsymbol{\Sigma}_{\hat{\omega}^{3D}}$, $E(\tau)$, $E(\alpha_t)$, and $E(\gamma)$ is required to obtain $\boldsymbol{\mu}_{\hat{\omega}^{3D}}$ and $\boldsymbol{\Sigma}_{\hat{\omega}^{3D}}$. The iteration continues until a convergence criterion is satisfied. For example, the relative difference between measurements $\boldsymbol{y}^{3D}$ and $\hat{\boldsymbol{y}}^{3D} = \mathbf{A}\boldsymbol{\mu}_{\hat{\omega}^{3D}}$, e.g., $r_E = \sum_{k=1}^{M} [y^{3D}(k) - \hat{y}^{3D}(k)]^2 / \sum_{k=1}^{M} [y^{3D}(k)]^2$ does not decrease with further iterations. $y^{3D}(k)$ and $\hat{y}^{3D}(k)$ represent the kth elements of $\boldsymbol{y}^{3D}$ and $\hat{\boldsymbol{\mu}}^{3D}$, respectively.

Subsequently, the 3D dataset $\hat{\underline{\mathbf{F}}}$ can be interpolated statistically, and the mean and interpolation standard deviation (SD) of each element are expressed as

$$
\mu_{\hat{\underline{F}}}(i, j, k) = \sum_{t=1}^{N} \mu_{\hat{\omega}_t^{3D}} \underline{\mathbf{B}}_t^{3D}(i, j, k) = (\boldsymbol{b}_{i,j,k}^T)\boldsymbol{\mu}_{\hat{\omega}^{3D}}
$$

$$
\sigma_{\hat{\underline{F}}}(i, j, k) = \sqrt{(\boldsymbol{b}_{i,j,k}^T)\boldsymbol{\Sigma}_{\hat{\omega}^{3D}}\boldsymbol{b}_{i,j,k}}, \tag{25}
$$

where $\mu_{\hat{F}}(i, j, k)$ and $\sigma_{\hat{F}}(i, j, k)$ respectively represent the mean (or expectation) and interpolation SD of element $\hat{\underline{F}}(i, j, k)$ and $\boldsymbol{b}_{i,j,k}$ is a column vector that records element $\underline{\mathbf{B}}_t^{3D}(i, j, k)$ in an increasing order of t. Equation (25) shows that given the measurement $\boldsymbol{y}^{3D}$, each element, particularly the measurements at unsampled locations of $\underline{\mathbf{F}}$, can be interpolated with the interpolation error given by $\sigma_{\hat{F}}(i, j, k)$. Furthermore, Eq. (25) shows that each element $\mu_{\hat{F}}(i, j, k)$ is a weighted summation of $\boldsymbol{\mu}_{\hat{\omega}^{3D}}$, which in turn is a function of the measurements $\boldsymbol{y}^{3D}$ [see Eq. (20)]. Similarly, $\sigma_{\hat{F}}(i, j, k)$ is a weighted summation of the elements $\boldsymbol{\Sigma}_{\hat{\omega}^{3D}}$, which is also a function of the measurements $\boldsymbol{y}^{3D}$. All of the elements in $\boldsymbol{y}^{3D}$ therefore contribute to $\mu_{\hat{F}}(i, j, k)$ and $\sigma_{\hat{F}}(i, j, k)$. Note also that neither the assumption that the 3D dataset is either isotropic or anisotropic nor the stationarity assumptions of the 3D dataset are required for the proposed method. This approach is thus robust and applicable when interpolating a 3D dataset with trends (i.e., so-called non-stationary data) or with anisotropic characteristics (e.g., anisotropic auto-correlation) (Wang et al. 2019).

5 Implementation Procedure

In this section, the implementation procedure of the proposed method is summarized as nine steps, which are discussed in detail below.

Step 1: Obtain sparse measurements and their locations in a 3D field and the size of the 3D field of interest. Consider, for example, a field with a length of $h_1 = 127$ m and width of $h_2 = 127$ m in the x- and y-directions, respectively, and a depth of $h_3 = 15$ m in the z-direction.

Step 2: Specify a resolution of interest for each direction, such as $\eta_1 = \eta_2 = \eta_3 = 1$ m for the x-, y- and z-directions, respectively, and calculate the dimensions of the 3D dataset to be interpolated as $N_1 = h_1/\eta_1 + 1 = 128$, $N_2 = h_2/\eta_2 + 1 = 128$, and $N_3 = h_3/\eta_3 + 1 = 16$.

Step 3: Construct the 3D basis function $\underline{\mathbf{B}}_t^{3D}$ (see "Appendix A"), then construct matrix $\mathbf{A}$ and $\boldsymbol{y}^{3D}$ following the procedure detailed in Sect. 3.

Step 4: Initialize all parameters. For example, set parameters $a_0 = b_0 = d_0 = 10^{-4}$ and $c_0 = 1$. Set the initial values as $E(\alpha_t) = E(\gamma) = 10^{-4}$ and $E(\tau) = 100/\text{var}(\boldsymbol{y}^{3D})$, where $\text{var}(\boldsymbol{y}^{3D})$ represents the measurement variance. The setting $E(\tau) = 100/\text{var}(\boldsymbol{y}^{3D})$ is suggested by Tipping (2001) and Ji et al. (2008) to ensure a good starting point during the iteration.

Step 5: Update $\boldsymbol{\mu}_{\hat{\omega}^{3D}}$ and $\boldsymbol{\Sigma}_{\hat{\omega}^{3D}}$ [see Eq. (20)] using $E(\alpha_t)$ and $E(\tau)$.

Step 6: Calculate a_t and b_t as $a_t = \mu^2_{\hat{\omega}^{3D}_t} + \sigma^2_{\hat{\omega}^{3D}_t}$, using the updated $\boldsymbol{\mu}_{\hat{\omega}^{3D}}$ and $\boldsymbol{\Sigma}_{\hat{\omega}^{3D}}$ and $b_t = E(\gamma)$, respectively, then update $E(\alpha_t)$ using Eq. (21a).

Step 7: Calculate c_n and d_n as $c_n = M/2 + c_0$ and $d_n = d_0 + 1/2[(\boldsymbol{y}^{3D})^T(\boldsymbol{y}^{3D}) - 2\boldsymbol{\mu}^T_{\hat{\omega}^{3D}}\mathbf{A}^T\boldsymbol{y}^{3D} + E[(\hat{\omega}^{3D})^T\mathbf{A}^T\mathbf{A}(\hat{\omega}^{3D})]]$, using the updated $\boldsymbol{\mu}_{\hat{\omega}^{3D}}$ and $\boldsymbol{\Sigma}_{\hat{\omega}^{3D}}$ together with Eq. (22), then update $E(\tau)$ using Eq. (21b).

Step 8: Calculate γ_a as $\gamma_a = N + a_0$ and update γ_b using $\gamma_b = b_0 + \sum_{t=1}^{N} E(\alpha_t^{-1})$, where $E(\alpha_t^{-1})$ is expressed in Eq. (24). $E(\gamma)$ is then updated as $E(\gamma) = \gamma_a/\gamma_b$ using Eq. (23).

Step 9: Return to Step 5 and continue the iteration until the relative difference between measurements $\boldsymbol{y}^{3D}$ and $\hat{\boldsymbol{y}}^{3D} = \mathbf{A}\boldsymbol{\mu}_{\hat{\omega}^{3D}}$ (i.e., $r_E = \sum_{k=1}^{M}[y^{3D}(k) - \hat{y}^{3D}(k)]^2 / \sum_{k=1}^{M}[y^{3D}(k)]^2$) is smaller than a prescribed small value (e.g., 10^{-3}) and r_E does not decrease with further iterations.

Note that although the derivation of the distributions $q(\hat{\omega}^{3D})$, $q(\alpha_t)$, $q(\gamma)$, $q(\tau)$, and their expectations seems tedious, the iteration process described above is relatively simple and straightforward. Furthermore, such an estimation procedure might lead to smoothed results when compared with simulation methods, as previously discussed in the literature (Yamamoto 2008). Note that the integration of the simulation methods (e.g., sequential Gaussian simulation) with the proposed method might be computationally extensive for high-dimensional datasets (e.g., 3D datasets in this study) due to the inverse of the estimated covariance structure, which has very high dimensions. In contrast, the proposed method is computationally efficient (i.e., much faster than simulations), particularly for high-dimension data. An estimation using the proposed method is therefore adopted in this study, and the development of the methods in this study with sequential Gaussian simulations will be explored in a future study to tackle the computational efficiency problem for 3D datasets. In the next section, the proposed method and procedure are demonstrated using a numerical example.

6 Numerical Example

In this section, the proposed method is illustrated using a simulated 3D dataset $\underline{\mathbf{F}}$ of a geo-quantity Q. The dataset $\underline{\mathbf{F}}$ represents the variations of Q in a 3D field, which has a length of 127 m over both the x- and y-directions and a depth of 15 m along the z-direction. In this example, $\underline{\mathbf{F}}$ has a resolution of 1 m in all directions. As a result, $\underline{\mathbf{F}}$ has $128 \times 128 \times 16 = 262{,}144$ data points in total, as shown in Fig. 1. Note that $\underline{\mathbf{F}}$ in Fig. 1 is simulated from a 3D Gaussian random field with a non-stationary mean of $\mu_Q = 30 - 0.001(x - 64)^2 - 0.0005(y - 64)^2 - 0.05(z - 8)^2$ and a constant SD $\sigma_Q = 3$, together with an anisotropic exponential auto-correlation function

$$\rho(\Delta x, \Delta y, \Delta z) = \exp\left(-2\sqrt{\frac{(\Delta x)^2}{\lambda_x^2} + \frac{(\Delta y)^2}{\lambda_y^2} + \frac{(\Delta z)^2}{\lambda_z^2}}\right), \qquad (26)$$

where $\Delta x = (x_i - x_m)$, $\Delta y = (y_j - y_n)$, and $\Delta z = (z_k - z_l)$ represent the relative distances between two points $\underline{\mathbf{F}}(i, j, k)$ and $\underline{\mathbf{F}}(m, n, l)$ in the x-, y- and z-directions, respectively, and λ_x, λ_y, and λ_z represent the correlation lengths within which the

geo-quantity Q is highly auto-correlated. In this example, λ_x, λ_y, and λ_z are taken as $\lambda_x = \lambda_y = 40$ m and $\lambda_z = 4$ m. The correlation lengths are consistent with those of geo-quantities reported in the literature (Phoon and Kulhawy 1999). Note that an exponential auto-correlation function is adopted here because it is often used in practice (Zhao and Wang 2018b; Liu et al. 2018). Equation (26) is used only for illustration and validation purposes (e.g., for simulating the 3D random field samples used for validating the proposed method). The proposed method also applies to other auto-correlation functions. Given the above parameters and the auto-correlation function in Eq. (26), $\underline{\mathbf{F}}$ in Fig. 1 is obtained using a random field generator [e.g., a circulant embedded algorithm by Dietrich and Newsam (1993) and Kroese and Botev (2015)]. This 3D dataset $\underline{\mathbf{F}}$ with $128 \times 128 \times 16 = 262{,}144$ data points in total is denoted as the original 3D data hereafter. In this example, $M = 100$ measurements $\mathbf{y}^{3D}$ are taken as the input. The $M = 100$ measurements are taken from 10 boreholes drilled along the z-direction, with 10 measurements obtained randomly with depth in each borehole. Figure 3 shows the locations of the $M = 100$ measurements as open circles. The measurements versus their locations are used as the input, and the original 3D dataset $\underline{\mathbf{F}}$ (i.e., Fig. 1) is used only for comparison and validation. Although the measurements used in this example are taken borehole by borehole to mimic the commonly used procedure in site investigation practice, this is not a requirement when using the proposed method. The approach developed here is equally applicable to measurements distributed in other patterns, e.g., at different depths. In addition, note that the $\underline{\mathbf{F}}$ data in Fig. 1, with dimensions of $128 \times 128 \times 16$, are used only for illustration. The proposed method is equally applicable to $\underline{\mathbf{F}}$ data with other dimensions.

Given the $M = 100$ data points measured from the 10 boreholes in Fig. 3, the high-resolution and spatially varied 3D quantity Q can be interpolated, as shown in Fig. 4(b). Figure 4(a) includes the original 3D dataset for comparison. Observe that the distribution of color in Fig. 4(b) is globally similar to that in Fig. 4(a). This indicates that the 3D dataset interpolated from the proposed method using sparse measurements is consistent with the original dataset. To further evaluate the interpolated 3D dataset, the population mean and SD are computed and compared with those of the original dataset in Table 1. Table 1 shows that the population mean and SD of the original 3D dataset are 26.88 and 3.59, respectively, and those of the interpolated data are 26.25 and 3.46, respectively. Table 1 also summarizes the absolute (relative) differences for the population means and SDs. The absolute difference between the means is 0.63 (with a relative difference of approximately 2.34%), and the absolute difference between the population SDs is 0.13 (with a relative difference of approximately 3.62%). The comparison suggests that the 3D dataset interpolated from the proposed method with 100 data points is reasonable.

To investigate further the results obtained using the proposed method, the residuals between the interpolated and original 3D dataset are computed at each location, as shown in Fig. 4(c). Figure 4(c) shows that although the absolute differences are relatively large in some locations, the magnitudes of most differences are generally small relative to those of the original 3D dataset. The small differences at most locations further demonstrate that the 3D dataset interpolated from the proposed method is reasonable and realistic. A close examination of Fig. 4(a) to (c) shows that some local variations of the original 3D dataset are not well reflected in the interpolated
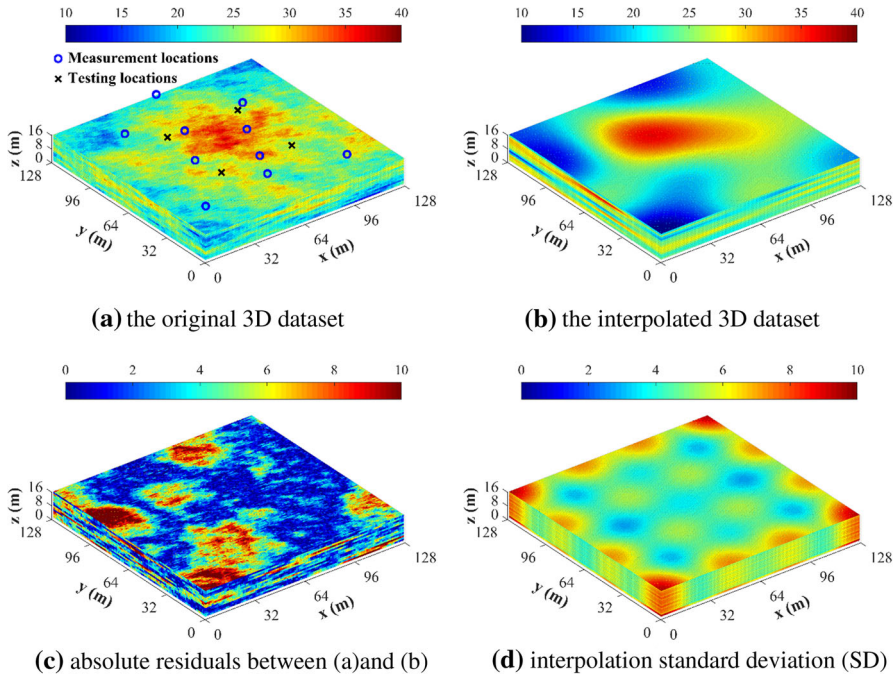
(a) the original 3D dataset

(b) the interpolated 3D dataset

(c) absolute residuals between (a)and (b)

(d) interpolation standard deviation (SD)

Fig. 4 Results from the proposed method using $M = 100$ measurements

Table 1 Population mean and standard deviation (SD) of the spatially varying 3D dataset	Statistics	Mean	SD
	Original 3D dataset with $128 \times 128 \times 16$ data points	26.88	3.59
	Interpolated 3D dataset from 100 data points	26.25	3.46
	Absolute difference	0.63	0.13
	Relative difference (%)	2.34	3.62

dataset. This is because the number of measurements used as the input in the proposed method is very small, and significant interpolation uncertainty has been introduced into the interpolation results. In the proposed method, this interpolation uncertainty is explicitly quantified. Figure 4(d) shows the interpolation SD associated with the results using Eq. (25). The magnitudes of the interpolation SD at most locations are slightly larger than or comparable to those of the absolute differences shown in Fig. 4(c). This implies a high probability that the true values of the original 3D dataset fall within the range defined by the interpolated 3D dataset ± 1.96 interpolation SD (i.e., $\mu_{\hat{F}}(i, j, k) \pm 1.96 \, \sigma_{\hat{F}}(i, j, k)$). For instance, approximately 98% of the elements in the original 3D dataset $\underline{\mathbf{F}}$ fall within the range defined by $\mu_{\hat{F}}(i, j, k) \pm 1.96 \, \sigma_{\hat{F}}(i, j, k)$ in this example. Hence, the quantified interpolation SD can be used as an indicator to evaluate the reliability of the obtained results. This is especially beneficial when the quantity values at unsampled locations must be estimated (Zhao et al. 2018b).
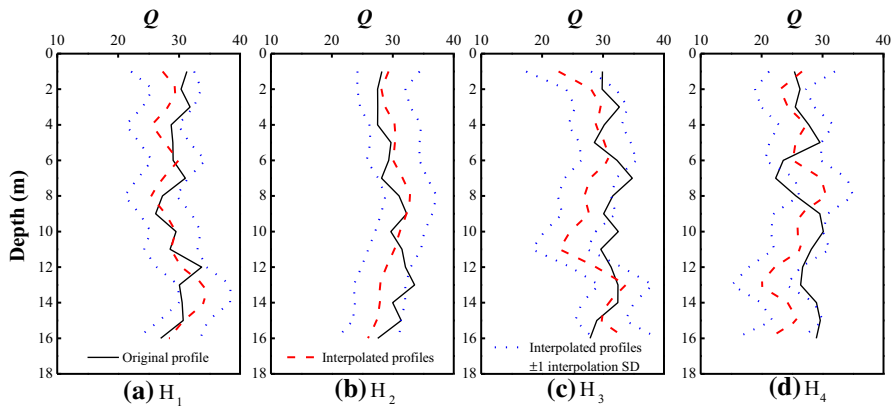
Fig. 5 Comparison between the original and interpolated profiles at boreholes H_1, H_2, H_3 and H_4

In addition to the above comparison, the variations of Q in four boreholes (H_1–H_4; Fig. 3) are extracted from the interpolated 3D dataset and compared with those from the original 3D dataset, as shown in Fig. 5(a) to (d), respectively. In Fig. 5, the dashed line represents the interpolated profile from the proposed method, whereas the solid line represents the original profile of Q at each corresponding borehole. Figure 5 also includes the quantified interpolation SD in terms of the interpolated profile $\pm$ interpolation SD, using a pair of dotted lines. Note that no measurements were taken along boreholes H_1–H_4. In Fig. 5, the dashed line is generally consistent with the solid line, and most local variations of the solid lines fall within the region defined by the two dotted lines. This agreement also indicates that the proposed method provides a reasonable interpolation for Q at the locations with no available measurements by using only a small number of measurements. Furthermore, a close examination of each subplot in Fig. 5 shows that the dashed lines in Fig. 5(a) to (b) (i.e., boreholes H_1 and H_2) are plotted slightly closer to the solid line than those in the other two subplots. This may be explained by noting that H_1 and H_2 are closer to the boreholes than H_3 and H_4, where measurements are available. The average distance between H_1 or H_2 and the three nearest boreholes where measurements are available is approximately 20.0 m or 20.8 m, respectively. In contrast, the average distance between H_3 or H_4 and the three nearest boreholes where measurements are obtained is approximately 25.6 m or 22.6 m, respectively, both of which are larger than both 20.0 m and 20.8 m.

Note that an interpolation by ordinary kriging (OK) is also performed for comparison using the same dataset, as shown by the open circles in Fig. 3. When OK is used for interpolation, the measurement data points should be stationary (e.g., without a trend). In such cases, de-trending should first be carried out on the measured data points, hoping that the residuals can be stationary. Suppose that the correct function form of the trend function is known, i.e., $\mu_Q = a_1 - a_2(x-a_3)^2 - a_4(y-a_5)^2 - a_6(z-a_7)^2$. Then, a_i ($i = 1, 2, \ldots, 7$) are obtained as 31.375, 0.002, 70.577, 0.001, 78.171, 0.036, and 8.480 using a least square approach, from which the trend function is plotted in Fig. 6(a). The parameters of the semi-variogram
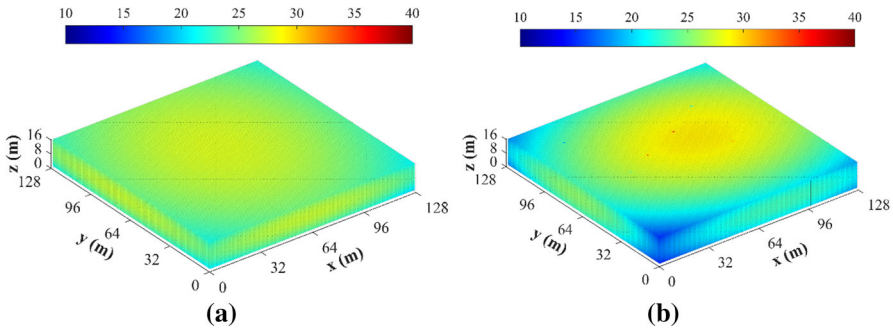
Fig. 6 a Trend function and **b** predicted mean from OK

$\rho(\Delta x, \Delta y, \Delta z) = \sigma^2 - \sigma^2 \exp\left(-2\sqrt{\frac{(\Delta x)^2}{\lambda_x^2} + \frac{(\Delta y)^2}{\lambda_y^2} + \frac{(\Delta z)^2}{\lambda_z^2}}\right)$ are then obtained as

$\sigma^2 = 5.39$, $\lambda_x = \lambda_y = 0.90$, and $\lambda_z = 3.52$ by minimizing the sum of the squared difference between the experimental semi-variogram of the residuals and the theoretical semi-variogram. To be consistent with the true semi-variogram used to simulate the 3D **F**, assumptions favorable to the OK interpolation are adopted so that $\lambda_x = \lambda_y$; the correct function form is known for the semi-variogram model and used herein. With these obtained parameters, the OK can then be carried out to obtain predictions at each location. Adding the trend function in Fig. 6(a) back to the mean from the OK yields the final results, as shown in Fig. 6(b). There are only slight differences between Fig. 6(a) to (b), particularly along the horizontal directions, indicating that Fig. 6(b) mainly reflects the result of the trend function. This is expected because the auto-correlation parameters $\lambda_x = \lambda_y = 0.90$ are severely underestimated in this example when compared with the $\lambda_x = \lambda_y = 40$ used in the simulation of the 3D **F**. This finding demonstrates that when the number of data points is small, there is a risk of severely underestimating the auto-correlation in the spatial direction (e.g., horizontal direction), even when the correct trend function form and semi-variogram model are already known. It is worth pointing out that selecting the most appropriate function form in 3D space from, say, 100 data points measured from 10 boreholes, is not trivial. Different function forms for the trend function might lead to different results (Fenton 1999a, b; Wang et al. 2019; Hu et al. 2020). The OK performance improves significantly with an increasing number of data points due to the proper estimation of the semi-variogram parameters. In contrast, de-trending is not required in the proposed method, and the 100 data points versus their locations are used directly as the input to produce the predictions, as discussed in detail in the previous section. The results from the proposed method are generally consistent with those in Fig. 1, although there are some local differences. The proposed method offers a supplement to kriging, especially when the number of measurements is small and kriging parameters cannot be reliably obtained.
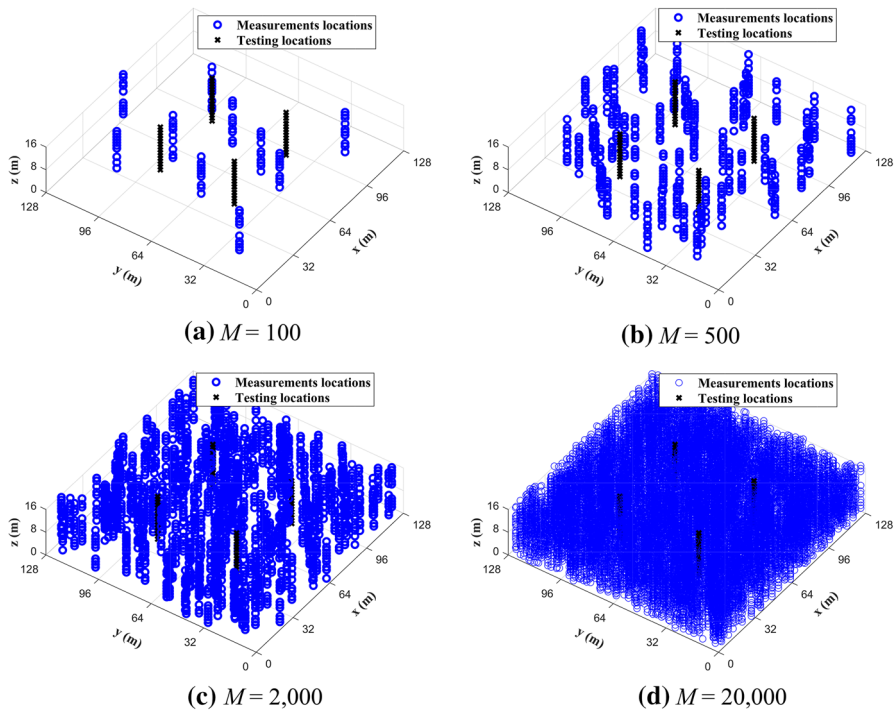
(a) $M = 100$

(b) $M = 500$

(c) $M = 2,000$

(d) $M = 20,000$

Fig. 7 Measurement locations and four boreholes under different M scenarios in 3D space

7 Effect of Number of Measurements

In this section, the proposed method is applied to scenarios with different numbers of measurements, e.g., $M = 500$, 2,000, and 20,000, in addition to the $M = 100$ example in the previous section. The $M = 100$, 500, 2000, and 20,000 measurements are approximately 0.04%, 0.19%, 0.76%, and 7.63%, respectively, of the total number of data points of $\underline{F}$ (262,144). Figure 7(a) to (d) show the measurement locations corresponding to $M = 100$, 500, 2,000, and 20,000, respectively. Each subplot of Fig. 7 also includes four boreholes, H_1–H_4, for later comparison. Note that in all of the M scenarios, no measurements were taken from boreholes H_1–H_4.

Given the measurements for each M scenario, the interpolated 3D dataset, absolute difference between the interpolated and original 3D datasets, and interpolation SD are obtained and shown in Figs. 8, 9 and 10, respectively, following the procedure detailed in Sect. 5 and illustrated in Sect. 6. Figure 8(a) to (d) show that the 3D datasets interpolated from the proposed method become increasingly similar to the original dataset with increasing M, and an increasing degree of the local variations of the original 3D dataset is reflected in the interpolation results. As a result, the absolute differences between the interpolated and original 3D datasets become increasingly small with increasing M, as shown in Fig. 9(a) to (d). When M is large (e.g., 20,000), nearly all of the residuals approach zero. These observations indicate that the results obtained from the proposed method become increasingly accurate with increasing
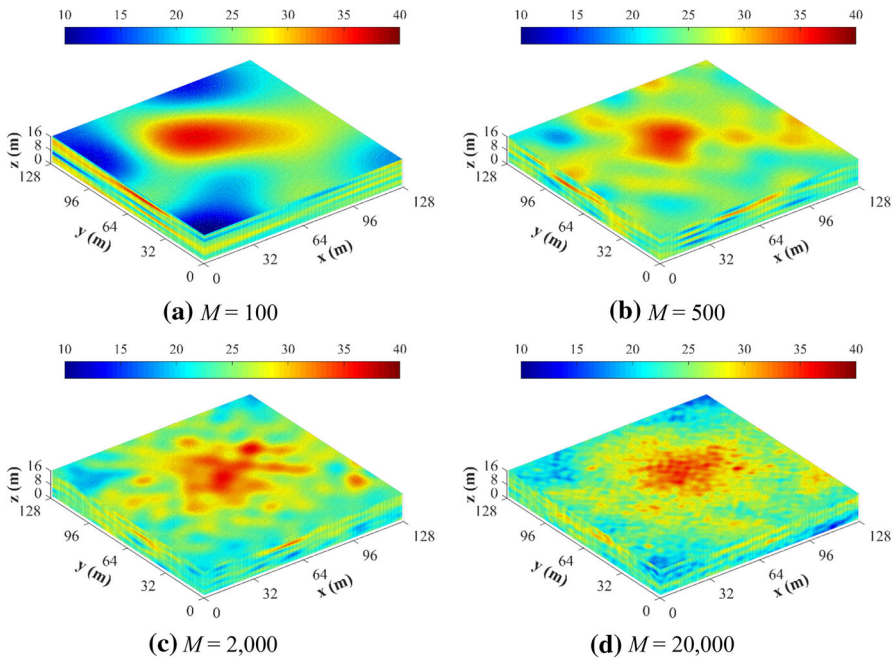
Fig. 8 Interpolated 3D dataset under different M scenarios

M. This is reasonable because more information about the original 3D dataset is available and used as the input with higher M. In addition, by respectively examining Figs. 9(a) to (d) and 10(a) to (d), the interpolation SD obtained from the proposed method is found to be slightly larger or comparable to the absolute residuals for each M scenario. This further implies that the true values of the original 3D dataset likely fall within the range defined by the interpolated 3D dataset ± 1.96 interpolation SD (i.e., $\mu_{\hat{F}}(i, j, k) \pm 1.96\sigma_{\hat{F}}(i, j, k)$) for each M scenario, even at locations where no data are available (Zhao et al. 2018b). Figure 10 shows that the interpolation SD at each location significantly decreases with increasing M. When M is large (e.g., 20,000), the SD obtained from the proposed method is reduced to a very small value. These observations suggest that the results from the proposed method become increasingly confident and reliable with increasing M.

Because it is difficult to visualize the variations of Q within the interpolated 3D dataset, the variations of Q from the four boreholes (H_1–H_4; Fig. 7) are extracted from the 3D interpolation results for each M scenario. Figure 10(a) to (d) summarize the interpolated Q profiles for boreholes H_1–H_4, respectively. In each subplot of Fig. 11, the interpolated Q profile is shown as open circles, open squares, open triangles, and crosses for the $M = 100$, 500, 2,000 and 20,000 scenarios, respectively. Figure 11 generally shows that the interpolated profiles gradually converge to the original profile with increasing M for each borehole. Consider, for example, the case of H_2. When M is small (e.g., 100), the interpolated profile (open circles) is plotted relatively far away from the original profile (i.e., bold solid line) (Fig. 11b). However, the interpolated
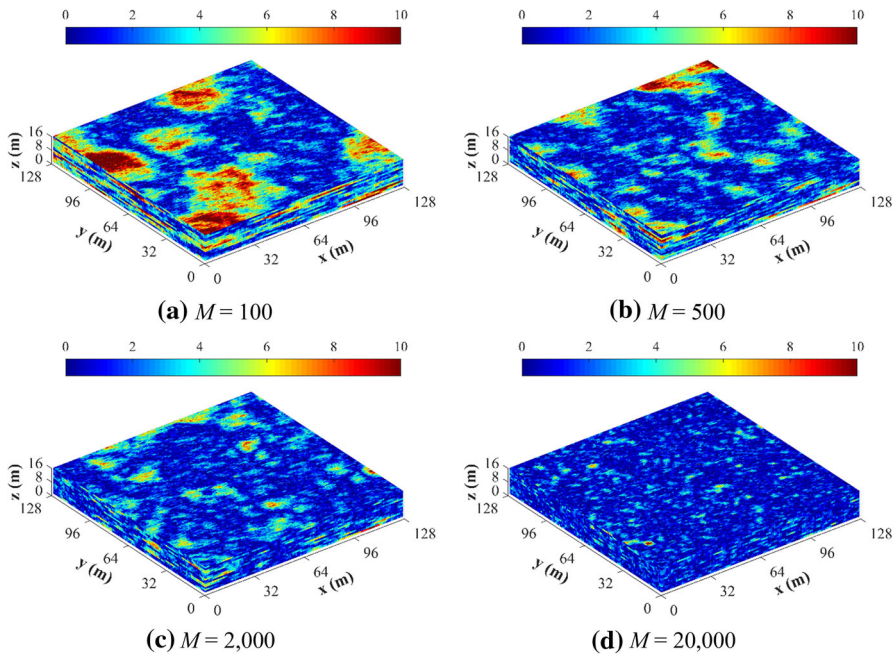
(a) $M = 100$ **(b)** $M = 500$

(c) $M = 2,000$ **(d)** $M = 20,000$

Fig. 9 Absolute difference between the original and the interpolated 3D dataset under different M scenarios

profile rapidly approaches the original profile with increasing M. When M is large (e.g., 20,000), the interpolated profile tends to overlap the original one (compare the crosses with the bold solid line in Fig. 11b). Similar results are observed in Fig. 11(a), (c), (d).

Figure 12(a) to (d) respectively plot the interpolation SDs at H_1 to H_4 for the four M scenarios, using the same symbols as those in Fig. 11. Similar to the observations in Fig. 10, the interpolation SD in each borehole is significantly reduced with increasing M. When M is large, the interpolation SD at each location reduces to approximately 1. This interpolation uncertainty is relatively small, compared with the mean, i.e., 26.88 of the original 3D dataset. A reduction of the interpolation uncertainty implies an increase in the reliability of the interpolation results from the proposed method with increasing M. The observations shown in Figs. 11 and 12 together provide further evidence that both the interpolated 3D dataset and quantified estimation uncertainty from the proposed method are reasonable.
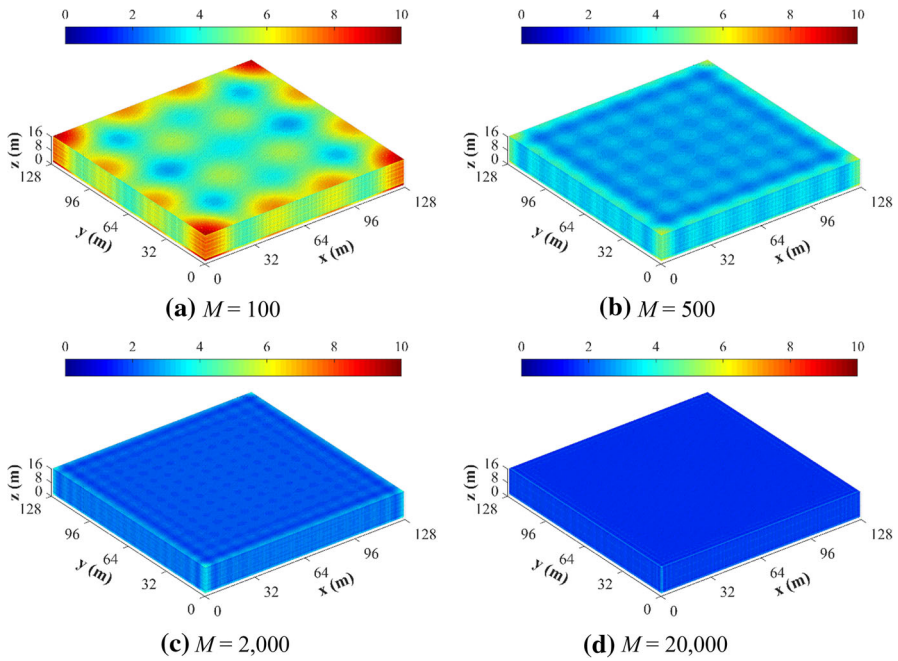
Fig. 10 Interpolation standard deviation (SD) associated with the 3D dataset under different *M* scenarios
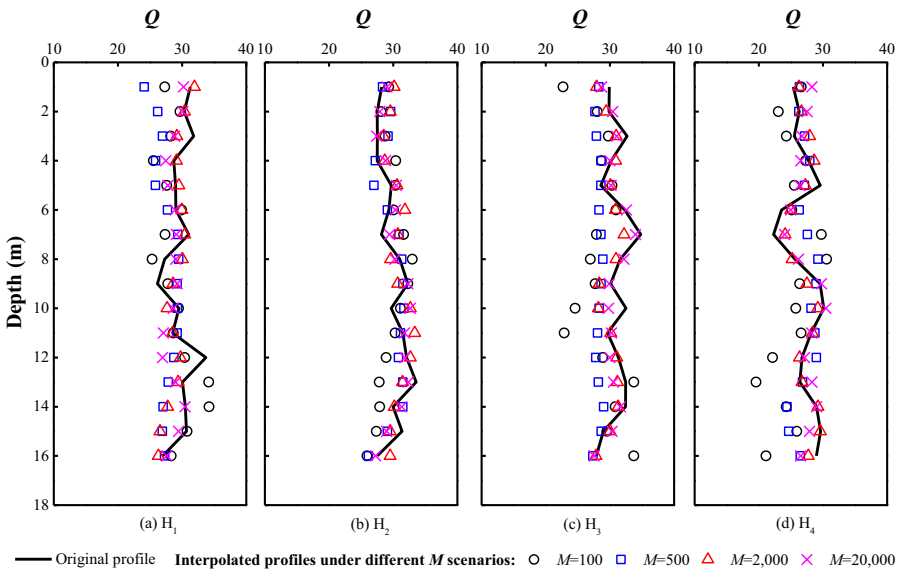

Fig. 11 Original and interpolated profiles at different boreholes H₁ to H₄
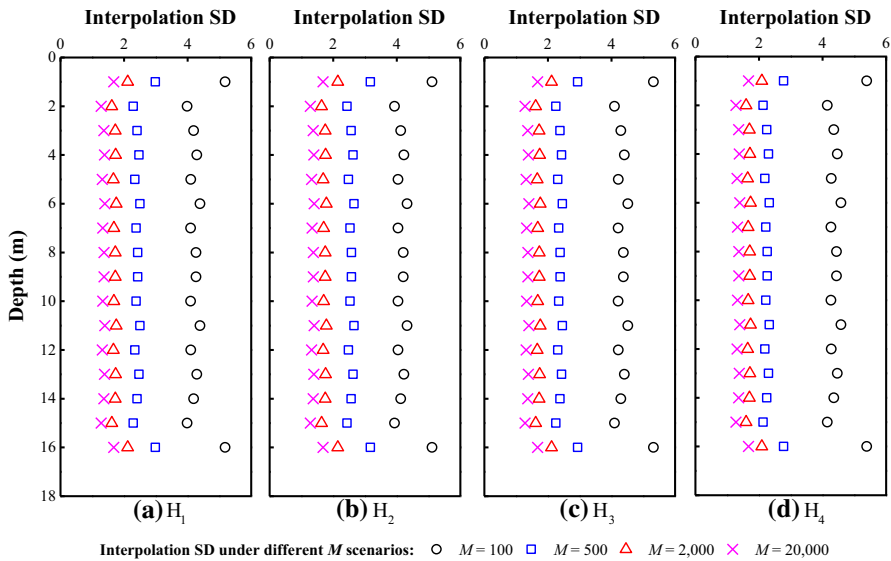
Fig. 12 Interpolation standard deviation (SD) at boreholes H_1 to H_4 under different M scenarios

8 Conclusions

A non-parametric method was proposed to statistically interpolate high-resolution and spatially varied 3D geo-data from sparse measurements. The proposed method systematically integrates compressive sensing with VBI, and uses sparse measurements and their 3D space locations as input to return a high-resolution 3D dataset together with the quantified uncertainty at each location. The interpolated 3D dataset and quantified uncertainty can be used further to achieve the desired design or analysis in geoscience. Note that the proposed method is data-driven and non-parametric, and an estimation of the correlation structure (e.g., semi-variogram) is not required. It may therefore be used as a supplement to powerful existing methods (e.g., kriging or GPR) for spatial interpolation, especially when the correlation structure among measurements cannot be reliably estimated. It is worth pointing out that the proposed method does not consider local variability during the interpolation, and sequential Gaussian simulations may be integrated with the proposed method when local variability is of great concern.

The equations are derived stepwise, and the procedure is illustrated using a series of numerical examples. The results show that the 3D dataset interpolated using the proposed method is consistent with the original dataset, and the quantified interpolation uncertainty is reasonable, even when the measurements are relatively sparse and limited. The interpolation uncertainty may be used as an indicator to evaluate the reliability of the interpolation results. The effect of the number of measurements is also investigated. The findings show that the interpolation results gradually converge to the original 3D dataset as the number of measurements increases, and the interpolation uncertainty is substantially reduced.

Appendix A: Construction of $\underline{\mathbf{B}}_t^{-3D}$

In this appendix, the 3D basis function $\underline{\mathbf{B}}_t^{3D}$ is reconstructed from columns of three independent 1D basis function matrices, i.e., $\mathbf{B}^1$, $\mathbf{B}^2$, and $\mathbf{B}^3$, which have dimensions of $N_1 \times N_1$, $N_2 \times N_2$, and $N_3 \times N_3$, respectively. The columns of $\mathbf{B}^1$, $\mathbf{B}^2$, and $\mathbf{B}^3$ represent orthonormal basis functions, e.g., discrete cosine functions. $\mathbf{B}^1$, $\mathbf{B}^2$, and $\mathbf{B}^3$ can be obtained using the formula for discrete cosine functions (see Salomon 2007) or constructed readily using a built-in function in commercial software, such as "dctmtx" in MATLAB (Mathworks 2020), which requires only N_1, N_2, and N_3 as the input. A 3D basis function is then constructed as $\underline{\mathbf{B}}_{i,j,k}^{3D} = \boldsymbol{b}_i^1 \otimes \boldsymbol{b}_j^2 \otimes \boldsymbol{b}_k^3$ ($i = 1, 2, \ldots, N_1$; $j = 1, 2, \ldots, N_2$; $k = 1, 2, \ldots, N_3$). The subscript of $\underline{\mathbf{B}}_{i,j,k}^{3D}$, i.e., "$i, j, k$" changes to "$t$" later in this paragraph. $\boldsymbol{b}_i^1$, $\boldsymbol{b}_j^2$, and $\boldsymbol{b}_k^3$ represent the ith, jth, and kth columns of $\mathbf{B}^1$, $\mathbf{B}^2$, and $\mathbf{B}^3$, respectively; "$\otimes$" represents an outer product and an element of $\underline{\mathbf{B}}_{i,j,k}^{3D}$, such as the element indexed by (m, n, l), i.e., $\underline{\mathbf{B}}_{i,j,k}^{3D}(m, n, l)$, is expressed as $\underline{\mathbf{B}}_{i,j,k}^{3D}(m, n, l) = b_i^1(m)b_j^2(n)b_k^3(l)$ (Kroonenberg 2008). $b_i^1(m)$, $b_j^2(n)$, and $b_k^3(l)$ are the mth, nth, and lth elements of $\boldsymbol{b}_i^1$, $\boldsymbol{b}_j^2$, and $\boldsymbol{b}_k^3$, respectively. After the construction of $\underline{\mathbf{B}}_{i,j,k}^{3D}$, the subscript "($i, j, k$)" of $\underline{\mathbf{B}}_{i,j,k}^{3D}$ changes to t for derivation convenience. "t" is numbered in increasing order of i, followed by j and k, respectively. It is worth noting that although the discrete cosine function is adopted in this study to construct the 3D basis function, other basis functions (e.g., wavelets functions) can also be used in the proposed method. The discrete cosine function is adopted here because it has analytical function forms and the basis function can be readily obtained.

Appendix B: Derivation of Eq. (15)

The expression of KL divergence defined in Eq. (14) is expanded to

$$
\begin{aligned}
KL(q||p) = &-\int q(\hat{\boldsymbol{\omega}}^{3D}, \boldsymbol{\alpha}, \gamma, \tau) \ln \frac{p(\hat{\boldsymbol{\omega}}^{3D}, \boldsymbol{\alpha}, \gamma, \tau | \boldsymbol{y}^{3D})}{q(\hat{\boldsymbol{\omega}}^{3D}, \boldsymbol{\alpha}, \gamma, \tau)} d\hat{\boldsymbol{\omega}}^{3D} d\boldsymbol{\alpha} d\gamma d\tau \\
= &-\int q(\hat{\boldsymbol{\omega}}^{3D}, \boldsymbol{\alpha}, \gamma, \tau)[\ln p(\hat{\boldsymbol{\omega}}^{3D}, \boldsymbol{\alpha}, \gamma, \tau | \boldsymbol{y}^{3D}) \\
&- \ln q(\hat{\boldsymbol{\omega}}^{3D}, \boldsymbol{\alpha}, \gamma, \tau)] d\hat{\boldsymbol{\omega}}^{3D} d\boldsymbol{\alpha} d\gamma d\tau.
\end{aligned}
\tag{27}
$$

In accordance with the rules of conditional probability, $p(\hat{\boldsymbol{\omega}}^{3D}, \boldsymbol{\alpha}, \gamma, \tau | \boldsymbol{y}^{3D}) = p(\hat{\boldsymbol{\omega}}^{3D}, \boldsymbol{\alpha}, \gamma, \tau, \boldsymbol{y}^{3D}) / p(\boldsymbol{y}^{3D})$. Substituting this expression into Eq. (27) and rearranging the terms lead to

$$
\begin{aligned}
KL(q||p) = - \int q(\hat{\boldsymbol{\omega}}^{3D}, \boldsymbol{\alpha}, \gamma, \tau) [\ln p(\hat{\boldsymbol{\omega}}^{3D}, \boldsymbol{\alpha}, \gamma, \tau, y^{3D}) \\
- \ln q(\hat{\boldsymbol{\omega}}^{3D}, \boldsymbol{\alpha}, \gamma, \tau) - \ln p(\boldsymbol{y}^{3D})] \mathrm{d}\hat{\boldsymbol{\omega}}^{3D} \mathrm{d}\boldsymbol{\alpha} \mathrm{d}\gamma \mathrm{d}\tau.
\end{aligned} \tag{28}
$$

Note that $\ln[p(\boldsymbol{y}^{3D})]$ is independent of the distribution $q(\hat{\boldsymbol{\omega}}^{3D}, \boldsymbol{\alpha}, \gamma, \tau)$. Therefore, $\int q(\hat{\boldsymbol{\omega}}^{3D}, \boldsymbol{\alpha}, \gamma, \tau) \ln p(\boldsymbol{y}^{3D}) \mathrm{d}\hat{\boldsymbol{\omega}}^{3D} \mathrm{d}\boldsymbol{\alpha} \mathrm{d}\gamma \mathrm{d}\tau = \ln p(\boldsymbol{y}^{3D})$. As a result, Eq. (28) is simplified as

$$
KL(q||p) = - \int q(\hat{\boldsymbol{\omega}}^{3D}, \boldsymbol{\alpha}, \gamma, \tau) \left[\ln \frac{p(\hat{\boldsymbol{\omega}}^{3D}, \boldsymbol{\alpha}, \gamma, \tau, \boldsymbol{y}^{3D})}{q(\hat{\boldsymbol{\omega}}^{3D}, \boldsymbol{\alpha}, \gamma, \tau)} \right] \mathrm{d}\hat{\boldsymbol{\omega}}^{3D} \mathrm{d}\boldsymbol{\alpha} \mathrm{d}\gamma \mathrm{d}\tau + \ln p(\boldsymbol{y}^{3D}).
\tag{29}
$$

Let $L(q) = \int q(\hat{\boldsymbol{\omega}}^{3D}, \boldsymbol{\alpha}, \gamma, \tau) \ln \left(\frac{q(\hat{\boldsymbol{\omega}}^{3D}, \boldsymbol{\alpha}, \gamma, \tau, \boldsymbol{y}^{3D})}{q(\hat{\boldsymbol{\omega}}^{3D}, \boldsymbol{\alpha}, \gamma, \tau)} \right) \mathrm{d}\hat{\boldsymbol{\omega}}^{3D} \mathrm{d}\boldsymbol{\alpha} \mathrm{d}\gamma \mathrm{d}\tau$. Equation (29) can then be rewritten as $KL(q||p) = -L(q) + \ln[p(\boldsymbol{y}^{3D})]$. Subsequently, $\ln[p(\boldsymbol{y}^{3D})] = L(q) + KL(q||p)$, i.e., Equation (15) can be obtained.

Appendix C: Derivation of $q(\hat{\boldsymbol{\omega}}^{3D})$, $q(\boldsymbol{\alpha})$, $q(\gamma)$, and $q(\tau)$

In this appendix, the framework for using VBI to derive the tractable distribution is introduced. Let $\boldsymbol{\Theta} = [\theta_1, \theta_2, \dots, \theta_n]^{\mathrm{T}}$ represent a set of random variables and "$p(\Theta|Data)$" represent the true posterior PDF of $\boldsymbol{\Theta}$ updated by "*Data.*" Suppose that "$p(\Theta|Data)$" has no analytical solution and VBI is adopted to seek an approximate distribution $q(\boldsymbol{\Theta})$ that can properly represent $p(\Theta|Data)$. As mentioned in the main text, $q(\boldsymbol{\Theta})$ is usually factorized as $q(\boldsymbol{\Theta}) = \prod_{i=1}^{n} q(\theta_i)$. In accordance with the derivation by Bishop (2006) (pp. 461–517), the $q(\theta_i)$ that minimizes the KL divergence between $q(\boldsymbol{\Theta})$ and $p(\Theta|Data)$ is then expressed as

$$
\begin{aligned}
\ln[q(\theta_i)] &= \int q(\boldsymbol{\Theta}_{-i}) \ln p(\boldsymbol{\Theta}, Data) \mathrm{d}\boldsymbol{\Theta}_{-i} + const \\
&= \int q(\boldsymbol{\Theta}_{-i}) \ln[p(Data|\boldsymbol{\Theta}) p(\boldsymbol{\Theta})] \mathrm{d}\boldsymbol{\Theta}_{-i} + const,
\end{aligned} \tag{30}
$$

where $\boldsymbol{\Theta}_{-i}$ represents $\boldsymbol{\Theta}$ with θ_i removed and "*const*" represents a term that ensures that the integration of $q(\theta_i) = 1$.

In this paper, the random variables of interest are $\hat{\boldsymbol{\omega}}^{3D}$, $\boldsymbol{\alpha}$, γ, and τ, and their approximate distribution can be individually derived using Eq. (30). Consider, for example,

$q(\hat{\boldsymbol{\omega}}^{3D})$. In accordance with Eq. (30), $q(\hat{\boldsymbol{\omega}}^{3D})$ is expressed as Eq. (17). Substituting Eqs. (6), (8–11), and (18) into Eq. (17) leads to

$$
\begin{aligned}
\ln[q(\hat{\boldsymbol{\omega}}^{3D})] &= \int q(\boldsymbol{\alpha})q(\gamma)q(\tau) \ln p(\hat{\boldsymbol{\omega}}^{3D}, \boldsymbol{\alpha}, \gamma, \tau, y^{3D}) d\boldsymbol{\alpha} d\gamma d\tau + const_1 \\
&= \int q(\tau) \ln p(\boldsymbol{y}^{3D}|\hat{\boldsymbol{\omega}}^{3D}, \tau) d\tau + \int q(\boldsymbol{\alpha}) \ln p(\hat{\boldsymbol{\omega}}^{3D}|\boldsymbol{\alpha}) d\boldsymbol{\alpha} + const_2,
\end{aligned}
\tag{31}
$$

where "$const_2$" represents the term that does not involve $\hat{\boldsymbol{\omega}}^{3D}$. Note that $q(\boldsymbol{\alpha})$ and $q(\gamma)$ are independent of $\ln p(\boldsymbol{y}^{3D}|\hat{\boldsymbol{\omega}}^{3D}, \tau)$. Therefore, $\int q(\boldsymbol{\alpha})q(\gamma)q(\tau) \ln p(\boldsymbol{y}^{3D}|\hat{\boldsymbol{\omega}}^{3D}, \tau) d\boldsymbol{\alpha} d\gamma d\tau = \int q(\tau) \ln p(\boldsymbol{y}^{3D}|\hat{\boldsymbol{\omega}}^{3D}, \tau) d\tau$. Similarly, $\int q(\boldsymbol{\alpha})q(\gamma)q(\tau) \ln p(\hat{\boldsymbol{\omega}}^{3D}|\boldsymbol{\alpha}) d\boldsymbol{\alpha} d\gamma d\tau = \int q(\boldsymbol{\alpha}) \ln p(\hat{\boldsymbol{\omega}}^{3D}|\boldsymbol{\alpha}) d\boldsymbol{\alpha}$. Subsequently, substituting Eqs. (6) and (8) into Eq. (31) and rearranging the terms lead to

$$
\begin{aligned}
\ln[q(\hat{\boldsymbol{\omega}}^{3D})] &= \int q(\tau) \ln p(\boldsymbol{y}^{3D}|\hat{\boldsymbol{\omega}}^{3D}, \tau) d\tau + \int q(\boldsymbol{\alpha}) \ln p(\hat{\boldsymbol{\omega}}^{3D}|\boldsymbol{\alpha}) d\boldsymbol{\alpha} + const_2 \\
&= -\frac{E(\tau)(\boldsymbol{y}^{3D} - \mathbf{A}\hat{\boldsymbol{\omega}}^{3D})^T(\boldsymbol{y}^{3D} - \mathbf{A}\hat{\boldsymbol{\omega}}^{3D})}{2} - \frac{(\hat{\boldsymbol{\omega}}^{3D})^T E(\mathbf{D}^\alpha)^T \hat{\boldsymbol{\omega}}^{3D}}{2} + const_3,
\end{aligned}
\tag{32}
$$

where "$const_3$" represents the terms that do not involve $\hat{\boldsymbol{\omega}}^{3D}$. Completing the square for $\hat{\boldsymbol{\omega}}^{3D}$ in Eq. (32) and rearranging the terms lead to

$$
\ln[q(\hat{\boldsymbol{\omega}}^{3D})] = -\frac{(\hat{\boldsymbol{\omega}}^{3D})^T[\mathbf{A}^T \mathbf{A} E(\tau) + E(\mathbf{D}^\alpha)]\hat{\boldsymbol{\omega}}^{3D} - 2(\hat{\boldsymbol{\omega}}^{3D})^T \mathbf{A}^T \boldsymbol{y}^{3D} E(\tau) + (\boldsymbol{y}^{3D})^T \boldsymbol{y}^{3D} E(\tau)}{2} + const_3.
\tag{33}
$$

Let $\boldsymbol{\Sigma}_{\hat{\omega}^{3D}} = [\mathbf{A}^T \mathbf{A} E(\tau) + E(\mathbf{D}^\alpha)]^{-1}$. Equation (33) is rewritten as

$$
\ln[q(\hat{\boldsymbol{\omega}}^{3D})] = -\frac{(\hat{\boldsymbol{\omega}}^{3D})^T (\boldsymbol{\Sigma}_{\hat{\omega}^{3D}})^{-1} \hat{\boldsymbol{\omega}}^{3D} - 2(\hat{\boldsymbol{\omega}}^{3D})^T (\boldsymbol{\Sigma}_{\hat{\omega}^{3D}})^{-1} \boldsymbol{\Sigma}_{\hat{\omega}^{3D}} \mathbf{A}^T \boldsymbol{y}^{3D} E(\tau)}{2} + const_4,
\tag{34}
$$

where "$const_4$" is a term that incorporates "$const_3$" and new terms that do not involve $\hat{\boldsymbol{\omega}}^{3D}$. Let $\boldsymbol{\mu}_{\hat{\omega}^{3D}} = \boldsymbol{\Sigma}_{\hat{\omega}^{3D}} \mathbf{A} \boldsymbol{y}^{3D} E(\tau)$. Equation (34) is rewritten as

$$
\begin{aligned}
\ln[q(\hat{\boldsymbol{\omega}}^{3D})] &= -\frac{(\hat{\boldsymbol{\omega}}^{3D})^T (\boldsymbol{\Sigma}_{\hat{\omega}^{3D}})^{-1} \hat{\boldsymbol{\omega}}^{3D} - 2(\hat{\boldsymbol{\omega}}^{3D})^T (\boldsymbol{\Sigma}_{\hat{\omega}^{3D}})^{-1} \boldsymbol{\mu}_{\hat{\omega}^{3D}} + (\boldsymbol{\mu}_{\hat{\omega}^{3D}})^T (\boldsymbol{\Sigma}_{\hat{\omega}^{3D}})^{-1} \boldsymbol{\mu}_{\hat{\omega}^{3D}}}{2} \\
&\quad + \frac{(\boldsymbol{\mu}_{\hat{\omega}^{3D}})^T (\boldsymbol{\Sigma}_{\hat{\omega}^{3D}})^{-1} \boldsymbol{\mu}_{\hat{\omega}^{3D}}}{2} + const_4 \\
&= \frac{(\hat{\boldsymbol{\omega}}^{3D} - \boldsymbol{\mu}_{\hat{\omega}^{3D}})^T (\boldsymbol{\Sigma}_{\hat{\omega}^{3D}})^{-1} (\hat{\boldsymbol{\omega}}^{3D} - \boldsymbol{\mu}_{\hat{\omega}^{3D}})}{2} + const_5,
\end{aligned}
\tag{35}
$$

where $\frac{(\boldsymbol{\mu}_{\hat{\boldsymbol{\omega}}^{3D}})^T(\boldsymbol{\Sigma}_{\hat{\boldsymbol{\omega}}^{3D}})^{-1}\boldsymbol{\mu}_{\hat{\boldsymbol{\omega}}^{3D}}}{2}$ is incorporated into the term "$const_5$". Therefore, $q(\hat{\boldsymbol{\omega}}^{3D})$ can be derived as

$$q(\hat{\boldsymbol{\omega}}^{3D}) = \frac{1}{\sqrt{(2\pi)^N \det(\boldsymbol{\Sigma}_{\hat{\boldsymbol{\omega}}^{3D}})}} \exp\left[-\frac{(\hat{\boldsymbol{\omega}}^{3D} - \boldsymbol{\mu}_{\hat{\boldsymbol{\omega}}^{3D}})^T(\boldsymbol{\Sigma}_{\hat{\boldsymbol{\omega}}^{3D}})^{-1}(\hat{\boldsymbol{\omega}}^{3D} - \boldsymbol{\mu}_{\hat{\boldsymbol{\omega}}^{3D}})}{2}\right]$$
$$\times \sqrt{(2\pi)^N \det(\boldsymbol{\Sigma}_{\hat{\boldsymbol{\omega}}^{3D}})} \exp(const_5), \tag{36}$$

A close examination of Eq. (36) shows that $\frac{1}{\sqrt{(2\pi)^N \det(\boldsymbol{\Sigma}_{\hat{\boldsymbol{\omega}}^{3D}})}}$
$\exp\left[-\frac{(\hat{\boldsymbol{\omega}}^{3D} - \boldsymbol{\mu}_{\hat{\boldsymbol{\omega}}^{3D}})^T(\boldsymbol{\Sigma}_{\hat{\boldsymbol{\omega}}^{3D}})^{-1}(\hat{\boldsymbol{\omega}}^{3D} - \boldsymbol{\mu}_{\hat{\boldsymbol{\omega}}^{3D}})}{2}\right]$ is the multivariate normal distribution, and its integration with respect to $\hat{\boldsymbol{\omega}}^{3D}$ is therefore 1. In addition, note that $q(\hat{\boldsymbol{\omega}}^{3D})$ is a PDF, which leads to $\int q(\hat{\boldsymbol{\omega}}^{3D}) d\hat{\boldsymbol{\omega}}^{3D} = 1$. As a result, the term "$\sqrt{(2\pi)^N \det(\boldsymbol{\Sigma}_{\hat{\boldsymbol{\omega}}^{3D}})} \exp(const_5)$" in Eq. (36) is equal to 1. In such a case, Eq. (36) is simplified as Eq. (19) in the main text.

Similarly, $q(\boldsymbol{\alpha})$ is derived as

$$q(\boldsymbol{\alpha}) = \prod_{t=1}^{N} \exp\left[-\frac{a_t\alpha_t + b_t\alpha_t^{-1}}{2}\right](\alpha_t)^{p-1} \times \frac{(a_t/b_t)^{p/2}}{2K_p(\sqrt{a_tb_t})} = \prod_{t=1}^{N} q(\alpha_t), \tag{37}$$

where $a_t = E[(\hat{\omega}_t^{3D})^2]$, $b_t = E(\gamma)$, $p = -1/2$, and $q(\alpha_t) = \exp\left[-\frac{a_t\alpha_t + b_t\alpha_t^{-1}}{2}\right](\alpha_t)^{p-1} \times \frac{(a_t/b_t)^{p/2}}{2K_p(\sqrt{a_tb_t})}$, which is a generalized inverse Gaussian (GIG) PDF (Zhao et al. 2015; Dumitru 2017). The mean or expectation of α_t is shown in Eq. (21a). In addition, note that $E(\alpha_t^{-1})$ is needed in the proposed method [see Eq. (24)], which cannot be directly evaluated even if $q(\alpha_t)$ is available. This is because $1/\alpha_t$ is a non-linear function of α_t. To address this problem, the PDF of $1/\alpha_t$, i.e., $q(\alpha_t^{-1})$ is derived as (Ang and Tang 2007)

$$q(\alpha_t^{-1}) = \frac{(a_t/b_t)^{p/2}}{2K_p(\sqrt{a_tb_t})} \exp\left[-\frac{a_t\alpha_t^{-1} + b_t\alpha_t}{2}\right](\alpha_t^{-1})^{p-1} \times \left|\frac{d(\alpha_t^{-1})}{\alpha_t}\right|$$
$$= \frac{(a_t/b_t)^{p/2}}{2K_p(\sqrt{a_tb_t})} \exp\left[-\frac{a_t\alpha_t^{-1} + b_t\alpha_t}{2}\right](\alpha_t^{-1})^{-p-1}. \tag{38}$$

Equation (38) shows that $(1/\alpha_t)$ follows a GIG with parameters b_t, a_t, and $-p$. The mean of $(1/\alpha_t)$, i.e., $E(\alpha_t^{-1})$, is obtained as Eq. (24).

In a manner similar to the derivation of $q(\hat{\boldsymbol{\omega}}^{3D})$ and $q(\boldsymbol{\alpha})$, both $q(\tau)$ and $q(\gamma)$ are derived to follow a Gamma distribution, which is shown in Eqs. (39) and (40), respectively

$$q(\tau) = \frac{(d_n)^{c_n}}{\Gamma(c_n)} \tau^{c_n - 1} \exp(-\tau d_n), \tag{39}$$

$$q(\gamma) = \frac{(\gamma_b)^{\gamma_a}}{\Gamma(\gamma_a)} \gamma^{(\gamma_a - 1)} \exp(-\gamma \gamma_b). \tag{40}$$

References

Ang AHS, Tang WH (2007) Probability concepts in engineering: emphasis on applications in civil and environmental engineering. Wiley, New York

Beal MJ (2003) Variational algorithms for approximate Bayesian inference. University of College, London

Bishop CM (2006) Pattern recognition and machine learning. Springer, New York, pp 461–517

Bishop CM, Tipping ME (2000) Variational relevance vector machines. In: Proceedings of the sixteenth conference on uncertainty in artificial intelligence, Morgan Kaufmann Publishers Inc., pp 46–53

Buland A, Kolbjørnsen O, Hauge R, Skjæveland Ø, Duffaut K (2008) Bayesian lithology and fluid prediction from seismic prestack data. Geophysics 73:C13–C21

Caiafa CF, Cichocki A (2013a) Computing sparse representations of multidimensional signals using Kronecker bases. Neural Comput 25:186–220

Caiafa CF, Cichocki A (2013b) Multidimensional compressed sensing and their applications. Wiley Interdiscip Rev Data Min Knowl Disco 3:355–380

Candès EJ, Romberg JK (2005) Signal recovery from random projections. In: SPIE international symposium on electronic imaging: computational imaging III, San Jose, California, pp 76–86

Candès EJ, Wakin MB (2008) An introduction to compressive sampling. IEEE Signal Proc Mag 25:21–30

Cao Z-J, Zheng S, Li D, Phoon K-K (2018) Bayesian identification of soil stratigraphy based on soil behaviour type index. Can Geotech J 56(4):570–586

Ching J, Phoon KK, Wu SH (2016) Impact of statistical uncertainty on geotechnical reliability estimation. J Eng Mech. https://doi.org/10.1061/(ASCE)EM.1943-7889.0001075

Cressie N (1993) Statistics for spatial data. Wiley, New York

Dietrich C, Newsam G (1993) A fast and exact method for multidimensional Gaussian stochastic simulations. Water Resour Res 29(8):2861–2869

Dimitrakopoulos R, Mustapha H, Gloaguen E (2010) High-order statistics of spatial random fields: exploring spatial cumulants for modeling complex non-Gaussian and non-linear phenomena. Math Geosci 42(1):65

Dumitru M (2017). Sparsity enforcing priors in inverse problems via normal variance mixtures: model selection, algorithms and applications. arXiv preprint arXiv:1705.10354

Fenton GA (1999a) Estimation for stochastic soil models. J Geotech Geoenviron Eng 125(6):470–485

Fenton GA (1999b) Random field modeling of CPT data. J Geotech Geoenviron Eng 125(6):486–498

Gilanifar M, Wang H, Sriram LMK, Ozguven EE, Arghandeh R (2019) Multi-task Bayesian spatiotemporal Gaussian processes for short-term load forecasting. IEEE Trans Ind Electron. https://doi.org/10.1109/TIE.2019.2928275

Griffiths DV, Marquez RM (2007) Three-dimensional slope stability analysis by elasto-plastic finite elements. Geotechnique 57(6):537–546

Hack R, Orlic B, Ozmutlu S, Zhu S, Rengers N (2006) Three and more dimensional modelling in geo-engineering. B Eng Geol Environ 65(2):143–153

Hillier MJ, Schetselaar EM, de Kemp EA, Perron G (2014) Three-dimensional modelling of geological surfaces using generalized interpolation with radial basis functions. Math Geosci 46(8):931–953

Hong Y, Wang L, Zhang J, Gao Z (2020) 3D elastoplastic model for fine-grained gassy soil considering the gas-dependent yield surface shape and stress-dilatancy. J Eng Mech 146(5):04020037

Horta A, Correia P, Pinheiro LM, Soares A (2013) Geostatistical data integration model for contamination assessment. Math Geosci 45:575–590

Hu Y, Wang Y, Zhao T, Phoon KK (2020) Bayesian supervised learning of site-specific geotechnical spatial variability from sparse measurements. ASCE-ASME J Risk Uncertain Eng Syst Part A Civ Eng 6(2):04020019

Huang Y, Beck JL, Wu S, Li H (2016) Bayesian compressive sensing for approximately sparse signals and application to structural health monitoring signals for data loss recovery. Probab Eng Mech 46:62–79

Huysmans M, Dassargues A (2009) Application of multiple-point geostatistics on modelling groundwater flow and transport in a cross-bedded aquifer (Belgium). Hydrogeol J 17(8):1901

Ji S, Xue Y, Carin L (2008) Bayesian compressive sensing. IEEE Trans Signal Process 56:2346–2356

Kroese DP, Botev ZI (2015) Spatial process simulation. In: Stochastic geometry, spatial statistics and random fields, Springer, pp 369–404

Kroonenberg PM (2008) Applied multiway data analysis. Wiley, Hoboken

Lamorey G, Jacobson E (1995) Estimation of semivariogram parameters and evaluation of the effects of data sparsity. Math Geol 27(3):327–358

Largueche FZB (2006) Estimating soil contamination with Kriging interpolation method. Am J Appl Sci 3(6):1894–1898

Li J, Heap AD (2011) A review of comparative studies of spatial interpolation methods in environmental sciences: performance and impact factors. Ecol Inform 6(3–4):228–241

Liu LL, Deng ZP, Zhang SH, Cheng YM (2018) Simplified framework for system reliability analysis of slopes in spatially variable soils. Eng Geol 239:330–343

Luo Z, Atamturktur S, Juang CH (2013) Bootstrapping for characterizing the effect of uncertainty in sample statistics for braced excavations. J Geotech Geoenviron Eng 139:13–23

Mariethoz G, Caers G (2015) Multiple-point geostatistics: stochastic modeling with training images. Wiley, London

Matheron G (1963) Principles of geostatistics. Econ Geol 58:1246–1266

Matheron G (1973) The intrinsic random functions and their applications. Adv Appl Probab 5(3):439–468

MathWorks I (2020) MATLAB: the language of technical computing. http://www.mathworks.com/products/matlab/. Accessed 6 May 2020

Moja SS, Asfaw ZG, Omre H (2018) Bayesian inversion in hidden Markov models with varying marginal proportions. Math Geosci 51(4):463–484

Murphy KP (2012) Machine learning: a probabilistic perspective. The MIT Press, London, pp 731–766

Petersen KB (2004) The matrix cookbook. Technical University of Denmark. http://www.cim.mcgill.ca/~dudek/417/Papers/matrixOperations.pdf. Accessed 13 Apr 2019

Phoon KK, Kulhawy FH (1999) Characterization of geotechnical variability. Can Geotech J 36(4):612–624

Rasmussen CE, Williams CKI (2006) Gaussian processes for machine learning. MIT Press, London

Remy N, Boucher A, Wu J (2009) Applied geostatistics with SGeMS: a user's guide. Cambridge University Press, Cambridge

Salomon D (2007) Data compression: the complete reference. Springer, New York

Schnabel U, Tietje O, Scholz RW (2004) Uncertainty assessment for management of soil contaminants with sparse data. Environ Manag 33(6):911–925

Shekaramiz M, Moon TK, Gunther JH (2017) Sparse Bayesian learning using variational Bayes inference based on a greedy criterion. In 2017 51st Asilomar conference on signals, systems, and computers, pp 858–862

Shekaramiz M, Moon TK, Gunther JH (2019) Exploration vs data refinement via multiple mobile sensors. Entropy 21(6):568

Shi C, Wang Y (2020) Non-parametric and data-driven interpolation of subsurface soil stratigraphy from limited data using multiple point statistics. Can Geotech J. https://doi.org/10.1139/cgj-2019-0843

Sivia D, Skilling J (2006) Data analysis: a Bayesian tutorial. OUP Oxford, New York

Strebelle S (2002) Conditional simulation of complex geological structures using multiple-point statistics. Math Geol 34(1):1–21

Tipping ME (2001) Sparse bayesian learning and the relevance vector machine. J Mach Learn Res 1:211–244

Wang Y, Zhao T (2017a) Bayesian assessment of site-specific performance of geotechnical design charts with unknown model uncertainty. Int J Numer Anal Methods Geomech 41:781–800

Wang Y, Zhao T (2017b) Statistical interpretation of soil property profiles from sparse data using Bayesian compressive sampling. Géotechnique 67:523–536

Wang Y, Huang K, Cao Z (2013) Probabilistic identification of underground soil stratification using cone penetration tests. Can Geotech J 50(7):766–776

Wang Y, Cao Z, Li D (2016) Bayesian perspective on geotechnical variability and site characterization. Eng Geol 203:117–125

Wang H, Wellmann JF, Li Z, Wang X, Liang RY (2017) A segmentation approach for stochastic geological modeling using hidden Markov random fields. Math Geosci 49(2):145–177

Wang X, Wang H, Liang RY, Zhu H, Di H (2018) A hidden Markov random field model-based approach for probabilistic site characterization using multiple cone penetration test data. Struct Saf 70:128–138

Wang Y, Zhao T, Hu Y, Phoon K-K (2019) Simulation of random fields with trend from sparse measurements without detrending. J Eng Mech 145:04018130

Williams CK (1998) Prediction with Gaussian processes: from linear regression to linear prediction and beyond. In: Learning in graphical models, Springer, Dordrecht, pp 599–621

Xiao T, Li DQ, Cao ZJ, Au SK, Phoon KK (2016) Three-dimensional slope reliability and risk assessment using auxiliary random finite element method. Comput Geotech 79:146–158

Yamamoto JK (2008) Estimation or simulation? That is the question. Comput Geosci 12(4):573–591

Yu L, Wei C, Jia J, Sun H (2016) Compressive sensing for cluster structured sparse signals: variational Bayes approach. IET Signal Process 10(7):770–779

Zhang T (2008) Incorporating geological conceptual models and interpretations into reservoir modeling using multiple-point geostatistics. Earth Sci Front 15(1):26–35

Zhang J, Zhang LM, Tang WH (2009) Bayesian framework for characterizing geotechnical model uncertainty. J Geotech Geoenviron Eng 135:932–940

Zhao T, Wang Y (2018a) Interpretation of pile lateral response from deflection measurement data: a compressive sampling-based method. Soils Found 58:957–971

Zhao T, Wang Y (2018b) Simulation of cross-correlated random field samples from sparse measurements using Bayesian compressive sensing. Mech Syst Signal Process 112:384–400

Zhao Q, Zhang L, Cichocki A (2015) Bayesian sparse Tucker models for dimension reduction and tensor completion. arXiv preprint arXiv:1505.02343

Zhao T, Hu Y, Wang Y (2018a) Statistical interpretation of spatially varying 2D geo-data from sparse measurements using Bayesian compressive sampling. Eng Geol 246:162–175

Zhao T, Montoya-Noguera S, Phoon KK, Wang Y (2018b) Interpolating spatially varying soil property values from sparse data for facilitating characteristic value selection. Can Geotech J 55(2):171–181